Helix Stability and Hydrophobicity in the Folding of Immunity Protein Im9

and

Kinetics of DNA-Binding to Transcription Factor GCN4

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)
vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich
von

Susanne Cranz
aus Deutschland

Promotionskomitee

Prof. Dr. Hans Rudolf Bosshard*
Prof. Dr. Sheena E. Radford
PD Dr. Ilian Jelesarov
*Vorsitz, Leitung der Dissertation

Zürich 2004

1. SUMMARY

The first part of this thesis describes a protein folding study focusing on the role of hydrophobicity and helicity in the folding pathway of the immunity protein Im9. The effect of variations in hydrophobicity and helical stability on the folding properties of structurally similar proteins is elucidated. The second part of the presented work describes the binding of GCN4 to its cognate DNA. The complex between a GCN4 dimer and DNA can be formed by two different pathways. We describe the equilibrium and kinetic properties of the two pathways and their relevance to DNA recognition.

Folding of Immunity Protein Im9

The native structure of a protein is encoded in its amino acid sequence; however it is not clear how structural information is encoded in a linear sequence and how the native state is attained. In the *hydrophobic collapse model* the protein adopts a conformation that exposes a minimum of hydrophobic surface and performs a search for its native state within the constraints of this compact conformation. In the *diffusion collision model* the collision of nascent secondary structure elements guides the folding process. Here the folding properties are influenced by the stability of the nascent secondary structure and the burial of hydrophobic surface area upon collision. In the *nucleation condensation model* a nucleus of structure initiates the folding process and as folding progresses secondary and tertiary structure stabilize concomitantly, radiating from the nucleus. The *unified model of folding* proposes that the folding mechanism is variable and determined by the ratio between the hydrophobicity and the intrinsic stability of secondary structure.

We have used the four-helix protein Im9 to test the influence of hydrophobicity and intrinsic stability of secondary structure on protein folding. We increased the helical propensity of three of the four helices and determined the folding properties of the mutants. The effect of stabilizing helices varies widely between the derivatives and is not correlated with the calculated increase in helical propensity. It appears that increasing helical propensity facilitates folding only as long as the unfolded state does not acquire helical structure. The effect of increasing the hydrophobicity of the solvent-exposed residues is also variable, suggesting that the effect of hydrophobicity depends on the role of the considered residue in folding. Comparison with another member of the immunity protein family, Im7, indicates that general

differences in helical propensity and hydrophobicity are not sufficient to explain the different folding properties that occur in this protein family.

DNA Recognition by Transcription Factor GCN4

Folded proteins exhibit a plethora of biological functions. Transcription factors for example bind DNA in a sequence specific way and control transcription. These molecules must avoid strong unspecific binding within the millions of basepairs that make up a genome and yet bind their cognate site with high affinity. In eukaryotes many transcription factors belong to the family of bZIP transcription factors. bZIP transcription factors consist of a basic region for DNA binding and a leucine zipper as dimerization domain. Although they are usually observed as dimers, there is mounting evidence that bZIP transcription factors can bind DNA not only as dimers but also as monomers, thus avoiding a situation where dimer formation becomes the rate limiting step. The association of a bZIP transcription factor and DNA can then be described as a thermodynamic cycle where thermodynamically equivalent monomer and dimer pathways lead to a ternary complex of a bZIP dimer associated with DNA.

We describe the DNA binding and dissociation of the monomer and the dimer of the bZIP transcription factor GCN4. The reactions of monomeric and dimeric GCN4 with DNA were studied in isolation using wildtype GCN4 and GCN4 derivatives under variable reaction conditions. We find that monomeric and dimeric GCN4 derivatives bind DNA at the same high rate and that different dissociation rates cause the differences in equilibrium affinity observed in equilibrium titrations. In the monomer pathway binding of the second monomer to the pre-formed GCN4 monomer-DNA complex is highly cooperative. With the help of the measured rate and equilibrium binding constants a model for a combined monomer-dimer pathway is calculated by numerical integration. According to this model the choice of the binding pathway is governed by the amount of pre-formed GCN4 dimer. In the cell nucleus, the local concentrations of GCN4 and DNA target site may decide on the pathway chosen.

ZUSAMMENFASSUNG

Der erste Teil dieser Arbeit beschreibt eine Proteinfaltungsstudie, welche die Rolle von Hydrophobizität und Helizität auf die Faltung des Immunitätsproteins Im9 erörtert. Der Effekt unterschiedlicher Hydrophobizität und Helixbildungstendenz auf den Faltungsmechanimus ähnlicher Proteine wird beschrieben. Der zweite Teil der Arbeit beschreibt die Bindung von GCN4 an seine DNA-Erkennungssequenz. Der Komplex zwischen einem GCN4 Dimer und der DNA kann sich auf zwei unterschiedlichen Wegen bilden. Wir untersuchen die thermodynamischen und kinetischen Parameter der beiden DNA-Bindungswege und deren Bedeutung für die DNA-Erkennung.

Faltung des Immunitätsproteins Im9

Die native Struktur eines Proteins ergibt sich aus der Aminosäurenequenz. Allerdings ist nicht klar, wie strukturelle Information in der linearen Abfolge der Aminosäuren enthalten ist und wie der native Zustand erreicht wird. Im *hydrophobic collaps model* erreicht das Protein zunächst eine Konformation, in der es ein Minimum hydrophober Oberfläche exponiert. Innerhalb der Vorgaben dieser kompakten Konformation entwickelt sich der native Zustand. Im *diffusion collision model* steuert die sich entwickelnde Sekundärstruktur den Faltungsprozess. Damit hängen die Faltungseigenschaften des Proteins von der Stabilität der sich entwickelnden Sekundärstruktur und der Reduktion der hydrophoben Oberfläche nach der Kollision ab. Im *nucleation condensation model* wird der Faltungsprozess durch einen Nukleus von Sekundärstruktur initiiert und Sekundär- und Tertiärstruktur entwickeln sich, ausgehend vom Nukleus, simultan. In einem *unified folding model* ist der Faltungsmechanismus variabel und von der Hydrophobizität der Aminosäurensequenz und deren Fähigkeit zur eigenständigen Entwicklung von Sekundärstruktur abhängig.

Wir beschreiben den Einfluss der Hydrophobizität und der Helixbildungstendenz auf die Faltung des vier-Helix Proteins Im9. Dazu erhöhten wir die Helixbildungstendenz von drei der vier Helices und charakterisierten die Mutanten. Der Effekt der erhöhten Helixbildungstendenz hängt davon ab, welche Helix betroffen ist, korreliert aber nicht mit dem Ausmass der vorhergesagten Zunahme der Helixstabilität. Eine hohe Helixbildungstendenz erleichtert die Proteinfaltung nur solange die Helices nicht bereits im ungefalteten Zustand strukturiert sind. Die Resultate zeigen ausserdem, dass die Unterschiede in der Faltungsreaktion der homologen Immunitätsproteine Im7 und Im9 durch Unterschiede in der Helixbildungstendenz und Hydrophobizität nicht vollständig erklärt werden können.

DNA Erkennung durch Transkriptionsfaktor GCN4

Gefaltete Proteine haben ausserordentlich vielfältige biologische Funktionen. Transkriptionsfaktoren, beispielsweise, binden sequenzspezifisch an DNA und kontrollieren die Transkription. Sie müssen starke unspezifische Assoziation innerhalb eines aus Millionen von Basenpaaren bestehenden Genoms vermeiden und nur an ihre Erkennungssequenz mit hoher Affinität binden. In Eukaryoten finden sich viele bZIP Transkriptionsfaktoren, welche aus einer basischen Region, die DNA bindet, und einem *leucine zipper* als Dimerisierungseinheit bestehen. Obwohl sie normalerweise als Dimer vorkommen, zeigt sich zunehmend, dass bZIP Transkriptionsfaktoren auch als Monomer an die DNA binden können, was einen kinetischen Engpass verhindert, bei dem die Dimerisierung zum geschwindigkeitsbegrenzenden Schritt wird. Die Bindung von GCN4 an DNA ist demnach ein thermodynamischer Zyklus bei dem die Bindung auf zwei energetisch gleichwertigen Wegen erfolgen kann, das heisst über einen Monomer-Weg oder einen Dimer-Weg.

Wir beschreiben die DNA Bindung des Dimers und des Monomers des bZIP Transkriptionsfaktor GCN4. Die Bindung von monomerem GCN4 und dimerem GCN4 an DNA konnte durch Verwendung von Wildtyp GCN4 und GCN4 Abkömmlingen und die Variation der Reaktionsbedingungen getrennt untersucht werden. Die monomeren GCN4 Derivate binden DNA mit der gleichen Geschwindigkeit wie dimeres GCN4, zeigen aber eine höhere Dissoziationskonstante, was durch die in Fluoreszenztitrationen gemessene niedrigere Affinität bestätigt wird. Im Monomer-Weg ist die Bindung eines zweiten GCN4 Monomers an einen bereits bestehenden GCN4 Monomer-DNA Komplex stark kooperativ. Mit den gemessenen Geschwindigkeits- und Gleichgewichtskonstanten wird ein Modell des kombinierten Monomer-Dimer-Wegs durch numerische Integration berechnet. Aufgrund dieses Modells bestimmt einzig die Menge des vorhandenen GCN4 Dimers den Bindungsweg. Im Zellkern dürften deshalb die lokalen Konzentrationen von GCN4 und DNA Bindungsstellen über den eingeschlagenen Bindungsweg entscheiden.

2. THE PROTEIN FOLDING PROBLEM

Proteins play an immense role for every aspect of biological life. Almost every chemical reaction in a biological system is catalyzed by proteins, which are termed enzymes. Some of the catalyzed reactions are simple chemical processes, which would occur, albeit immeasurably slowly, even without enzymes. Other reactions are extremely complex and involve not only several reactants but also numerous enzymes. Proteins also play a hugely important role in the intriguing process of growth and differentiation. Growth and differentiation are regulated by differential expression of genes in different tissues and coordination of different organs by hormones. Proteins mediate all of these processes. The function of every single organ or system of organs, from the nervous to the digestive or immune system is based on the structure and function of proteins. Besides these prominent functions, proteins serve as transporter molecules for smaller metabolites and are the molecular basis of coordinated movement which requires mechanical resistance and stability that is also conferred by proteins (*Stryer, 1990*).

Every naturally occurring protein is made up of a linear sequence of building blocks called amino acids. The sequence of amino acids that makes up a protein is called its primary structure. There are 20 biogenic amino acids; assuming an average protein length of 200 amino acids there can be 20^{200} different protein sequences. Only a minute fraction of this vast sequence space is actually populated by real proteins. Nevertheless: If there are 10 million species on earth and each of those have an average of 5000 genes we are left with 5×10^{10} sequences (*Koonin et al., 2002*).

To fulfill their biological functions the linear sequence of amino acids making up a protein must adopt a three-dimensional structure referred to as tertiary structure. Usually this occurs spontaneously in a process called folding suggesting that the three-dimensional tertiary structure is encoded in the primary structure. Several complex questions are raised by this observation (*Pain, 1994*):

►How is the three-dimensional tertiary structure encoded in the linear primary structure?

►What is the thermodynamic basis of the stability of the folded state relative to all other states?

►By which kinetic process or pathway does a protein adopt its native structure?

These three questions sum up what is called the "protein-folding problem". The protein-folding problem has far reaching consequences for medicine and biotechnology. If the three-dimensional structure of a protein and its stability could be predicted, proteins could be custom made to fulfill specific functions. If pathways leading to incorrectly folded proteins were fully understood, small molecules, drugs, could be developed to disrupt these pathways. This could lead to cures for cystic fibrosis and Alzheimer's disease as well as for other less common and less well-known diseases.

2.1. Models for Protein Folding

The central problem of protein folding was first addressed by C. Levinthal in 1968 *(Levinthal, 1968)* and is hence referred to as the "Levinthal paradox". Levinthal argued that a protein would never fold on a biologically relevant timescale if it randomly sampled the entire conformational space open to it. It follows that an unfolded protein must perform a directed rather than an unbiased search for its native state (*Fersht, 1998*). The scientific challenge is not so much resolving the Levinthal paradox but describing the mechanism which directs the search for the native conformation (*Honig, 1999*). At least three different mechanisms by which a protein can attain its native state have been proposed.

2.1.1. Hydrophobic Collapse Model

Several observations indicate the supreme importance of the hydrophobic effect for the folded protein. Firstly, non-polar solvents denature protein because they preferentially stabilize the unfolded state by solvating its exposed hydrophobic amino acid residues (*Bennion and Daggett, 2003*). Secondly, proteins show an unexpected cold destabilization. The physical basis of this observation is that non-polar solutes become more soluble in water at low temperatures (*Privalov and Gill, 1988*). Thirdly, inspection of three-dimensional structures of proteins as NMR and x-ray crystallography can obtain them, reveals that most of the hydrophobic residues of a polypeptide chain are found in the core of the protein, shielded from the aqueous solvent. Thus it largely is the hydrophobic effect that stabilizes the native structure of proteins relative to the various unfolded conformations (*Dill, 1990*).

In the hydrophobic-collapse model of protein folding a protein rapidly collapses around its hydrophobic side-chains giving rise to an intermediate structure (*Dill, 1985*). The volume and the conformational space occupied by the collapsed chain limit the ensuing search for the native conformation. Interestingly the population of partially folded intermediates under equilibrium conditions can be correlated to the charge/hydrophobicity ration of a sequence (*Uversky, 2002*). Measurements with sufficient time resolution should thus reveal the formation of a compact globule with a hydrophobic core but without optically active secondary structure. Although this has been observed in the folding of barstar, most other observable intermediates display varying amounts of secondary structure (*Agashe et al., 1995*). However, the recent advancement of methods which allow the folding reaction to be observed under physiological conditions with very high time resolution may lead to a reappraisal of the role of hydrophobic collapse in the very early stages of folding (*Sadqi et al., 2003*).

It has been suggested that the hydrophobic effect is not only important for the thermodynamic stability of the native state but may direct the folding reaction (*Akiyama et al., 2002*). Several studies indicate that specifically positioned hydrophobic residues can accelerate folding by guiding formation of tertiary structure (*Colon et al., 1996; Khorasanizadeh et al., 1996*), stabilizing the transition state or providing alternative folding pathways (*Calloni et al., 2003*). Although these effects indicate the importance of the hydrophobic effect for folding they do not provide direct support for the hydrophobic collapse model.

Theoretical studies about the role of hydrophobic interactions in the folding reaction reach the conclusion that hydrophobic collapse and formation of backbone hydrogen bonds, i.e. secondary structure must be coupled to be energetically feasible (*Fernandez et al., 2003*). The calculations indicate that hydrophobic collapse alone, without concomitant formation of secondary structure is not conducive to folding. Although compaction decreases the number of conformations available to the protein chain, thereby reducing the entropic barrier, it also prohibits some conformations that represent a low-barrier pathway. Essentially the entropic barriers to folding that were pointed out by Levinthal are replaced with equally insurmountable energetic barriers (*Gutin et al., 1995*).

2.1.2. Diffusion-Collision Model

In 1976 Karplus and Weaver (*Karplus and Weaver, 1976*) proposed the diffusion-collision model as a mechanism of folding. The diffusion-collision model is a hierarchic model of folding which separates the folding reaction into two separate and more or less resolvable processes: formation of local structure and interaction of these structures with one another (*Baldwin and Rose, 1999i; Baldwin and Rose, 1999ii*). In the first stage of folding elements of incipient secondary structure and hydrophobic clusters become populated. These very unstable elements are called microdomains. The microdomains diffuse until they collide with another microdomain. If the collision is productive, it gives rise to a larger and more stable element of structure consisting of two or more microdomains. When these higher-order microdomains contain native pairings of microdomains they are generally on a pathway leading to the native state. During the folding non-native pairings can occur, resulting in off-pathway or misfolded intermediates. However, intermediates do not necessarily represent a futile journey through non-native areas of the conformational space but may be a productive species found on the pathway to the native state. Like the model of hydrophobic collapse, the diffusion-collision model can also explain highly cooperative and apparently two-state manner folding transitions (*Karplus and Weaver, 1994*). The Levinthal paradox is resolved in the diffusion collision model by the introduction of microdomains, which provide a folding pathway that avoids a search of the entire conformational space available to the polypeptide chain. Instead the conformational search happens on the level of the microdomains, which are small enough to perform this search in a biologically accessible time-scale. The shortness of the microdomains makes them on one hand accessible to a random search mechanism and is on the other hand a cause for their only marginal stability.

The diffusion-collision model has been used to interpret the folding kinetics of all-helical proteins, because the helices themselves are obvious microdomains (*Islam et al., 2002*). In the diffusion-collision model the stability of the helical microdomains, the length of the chain segments between microdomains, the extent of solvent accessible area loss upon collision and the native structure are be used to make quantitative predictions regarding the folding rate and the formation of native and non-native intermediates (*Islam et al., 2002*). For the formation of non-native kinetic intermediates, which may slow folding, the loss of hydrophobic solvent accessible area upon cohesion of two microdomains is of special interest (*Capaldi et al., 2002*). A low ratio of native pairings to possible pairings also increases the possibility of transient non-native pairings (*Beck et al., 2001*). Experimental and theoretical studies indicate that in

proteins that feature intrinsically stable secondary structure and a high bias gap, which characterizes the stability of native relative to non-native contacts, the folding can be described by the diffusion-collision model (*Mayor et al., 2003; Zhou and Karplus, 1999; Zhu et al., 2003*).

2.1.3. Nucleation-Condensation Model

The possibility of characterizing the transient transition state by the protein engineering method (see chapter 2: Methods in Protein Folding Kinetics) introduced new perspectives to the protein-folding field (*Fersht et al., 1992; Matouschek et al., 1992; Serrano et al., 1992*). Applying this method to the folding of chymotrypsin inhibitor 2 (CI2) (*Itzhaki et al., 1995*), cold shock protein B (CSP B) (*Perl et al., 1998*), Barstar (*Nolting et al., 1997*) and other small two-state folders has led to a new model for protein folding. In the transition state, the whole amino acid chain shows a variety of weak interactions that center on a more stable but diffuse nucleus of native structure. Presumably, the physiologically denatured state in these proteins contains a nucleus of weak and transient native or nearly native structure, while the rest of the protein is disordered. During the folding reaction, the rest of the protein condenses around this nucleus that in turn is consolidated by the interactions resulting from this condensation. In the nucleation condensation model, therefore secondary and tertiary interactions form concurrently. In larger multidomain proteins the individual domains also fold in this manner and are then assembled in a diffusion-collision-like process (*Matouschek et al., 1992*). Analysis of a number of proteins shows that the consolidated region of structure in the transition state, meaning the nucleus, is usually a region in which secondary structure dominates (*Nolting and Andert, 2000*). Intermediates that have predominantly secondary structure in the presence of distorted, weak or fluctuating tertiary interactions are observed in numerous proteins, especially those showing a significant amount of helical secondary structure (*Eliezer et al., 1998; Ferguson et al., 1999; Matouschek et al., 1992*). Cooperative folding and unfolding and sensitivity to truncation of buried hydrophobic side chains indicate that these intermediates are not merely ensembles of transient fluctuating secondary structure (*Baldwin and Rose, 1999; Capaldi et al., 2002*). In the course of further analysis of non-two state folders it became apparent that hierarchic models of protein folding and the nucleation condensation mechanism have to be reconciled to obtain a universal model of protein folding (*Daggett and Fersht, 2003; Gianni et al., 2003*).

2.2. The Role of Secondary Structure in Protein Folding

Understanding the formation of secondary structure stabilized by local interactions, i.e. α-helices and β-turns and its role in the global folding process is of central importance. Studies carried out using computer lattices have produced contradictory results. A study by Moult and Unger using a Monte Carlo simulation method shows that foldability, i.e. ability to find the global energy minimum depends on the locality of the interactions. Sequences that include many stabilizing local interactions have a higher likelihood of folding. Small substructures that are local in sequence can form without a high entropic penalty and drive the subsequent collapse of the chain (*Unger and Moult, 1996*). Other studies show the opposite, namely that a large contribution of local interactions to the stability of the native state slows the rate of protein folding. This is because local interactions do not show a strong specificity for the native state but also stabilize other conformations, leading to a non-cooperative folding reaction. The resulting sequences do not show a pronounced energy minimum, meaning that they are thermodynamically stable only at very low temperatures at which their native state is kinetically inaccessible (*Abkevich et al., 1995; Govindarajan and Goldstein, 1995; Unger and Moult, 1996*).

Experimentally the role of secondary structure has been analyzed by increasing the stability of secondary structure elements through site-directed mutagenesis and by analyzing the effects of organic solvents that selectively stabilize α-helical conformations. These studies largely conclude that stabilization of helices results in the stabilization of the protein. Studies on ADA2h have shown that this stabilization can be due to the acceleration of the folding or deceleration of unfolding. The former case has been interpreted to mean that the affected helix is formed in the transition state, the latter indicates that the helix is fully formed only after the main transition state has been traversed (*Viguera et al., 1997*). This means that those states in which the helix is already structured are stabilized relative to those states in which the helix is not yet structured. A similar study conducted on CheY points out that the increase in stability of the protein is much smaller than the stabilization of the helices as measured in isolated peptides, indicating that not only the native state but also other conformations have been stabilized by the mutagenesis. The authors also find that increasing the contribution of local interactions compacts the denatured state (*Munoz et al., 1996*). This may mean that some helical structure is already present in the denatured state, which is therefore also stabilized. In this case, because the stability is measured as a relative number, i.e. ΔG_{ij}, no stabilization of

folded states relative to the unfolded state will be apparent. A study that investigates the role of helical propensity in the folding of acylphophatase highlights that effects resulting from stabilization of helical structure by mutagenesis are essentially identical to those that are achieved by using TFE, which increases stability of helical structure chemically (*Taddei et al., 2000*). The maximal increase in folding rate is achieved at a TFE concentration in which the denatured state has a similar amount of secondary structure as the native state (*Chiti et al., 1999*). Work published on the folding of BdpA, λ-repressor and GCN4-p1 dimerization confirms that helical propensity of a sequence, that is the stability of helices in the unfolded state, has to be taken into account if the diffusion-collision model is to make quantitative predictions of folding rates (*Burton et al., 1998; Myers and Oas, 1999; Myers and Oas, 2001*).

In 1998 Plaxco et al. (*Plaxco et al., 1998*) examined folding rate and transition state placement of single domain proteins under the aspect of a new parameter that they named contact order. Relative contact order is defined as the average sequence separation between all contacting residue pairs and is indicative of the protein's topology. A clear correlation between contact order and folding rates can be observed. Work recently conducted by Gong et al. indicates that contact order may be a proxy for fraction of stabilizing contacts that are local (*Gong et al., 2003*). They find that the total local secondary structure content including α-helices, tight turns, and β-hairpins is an equally good determinant for folding rates. The clear advantage of this theory is that local secondary structure, unlike contact order has a direct physico-chemical interpretation (*Srinivasan and Rose, 1999*).

2.3. The Role of Intermediates in Protein Folding

The characterization of intermediates and their role in the folding reaction are a central topic in the study of protein folding because folding through partially structured intermediates is one way of solving the Levinthal paradox.

Different models for protein folding assign different roles to intermediates. For the hydrophobic-collapse model and the diffusion-collision model they are important steps on the way to the native state although they may be hidden (*Karplus and Weaver, 1994*). In the nucleation-condensation model however, they should occur only in the folding of multidomain proteins at a stage where individual domains are formed but not yet docked together, as observed in barnase (*Matouschek et al., 1992*). Here intermediates that occur in the folding of

small single domain proteins are misfolded dead-end conformations that essentially slow the folding process. They represent a futile journey through non-native conformational space and do not occur when the folding process is optimized (*Schindler et al., 1995*).

The study of intermediates presents a number of challenges. An intermediate is only observable if it occupies an energy-well that is lower than any of the previous energy wells and encounters a barrier that is higher than all previous barriers; otherwise it is hidden (*Bai, 2003*). Even when intermediates are observable kinetically they are often thermodynamically unstable, which makes their characterization by x-ray crystallography or NMR impossible (*Fersht, 1993*). A number of intermediates that occur after the rate-limiting transition state in cytochrome c folding and are therefore not detectable in conventional kinetic experiments because they do not encounter a rate limiting barrier were identified by Bai et al. using a native state hydrogen exchange method (*Bai et al., 1995*). Later it was shown that the intermediates are on the major pathway for unfolding (*Bai, 1999; Hoang et al., 2002*). Similar techniques were employed to identify and characterize intermediates in the folding of the redesigned four helix bundle protein Rd-Apocyt b_{562} (*Chu et al., 1999; Takei et al., 2002*).

In many cases the formation of intermediates is evident already from the comparison of the stabilities and denaturant sensitivities obtained from conventional kinetic and thermodynamic measurements. If the two-state model fully describes the thermodynamic and kinetic behavior of the protein, the stability and denaturant sensitivity obtained by the two methods should agree (*Jackson and Fersht, 1991*). If the intermediate is stable enough under the experimental conditions, kinetic measurements alone can reveal folding through an intermediate. The most straightforward kinetic observation in this case is that the trace generated by folding of the protein cannot be described by an equation for a single exponential decay (*Jackson and Fersht, 1991*). To fully describe the trace generated by a folding reaction that encounters two or more rate limiting steps an exponential decay of the respective order must be used. When the formation of the intermediate is too rapid to be observed directly, a non-linear relationship between denaturant concentration and activation free energy called rollover and so-called burst phase amplitudes can be observed. This occurs when a folding intermediate with spectroscopic properties that are different from the denatured state accumulates within the dead-time of a stopped-flow measurement. When the accumulating intermediate has the same spectroscopic properties as the denatured state no burst phase amplitude can be observed (see chapter 3, Analysis of Kinetic Measurements of Protein Folding). A rollover occurring without burst

phase amplitude can also be attributed to a switch of the rate-limiting transition state at a certain denaturant concentration, to structural changes in the transition state according to the Hammond-postulate or to non-linear denaturant activity. However, a detailed study of 23 small proteins that show a non-linear relationship between denaturant concentration and activation free energy reveals that this behavior is best explained by sequential barriers (*Sanchez and Kiefhaber, 2003*). The structure of unstable intermediates is observable through the combination of the protein engineering method and kinetic methods with a very high time-resolution (*Fersht, 1993*). Another strategy for the structural characterization of intermediates is stabilizing them to a point where they can be observed at equilibrium and characterized by various spectroscopic methods. Especially when the structure of the native state and the intermediate are known mutations that destabilize the former while stabilizing the latter can be introduced (Graham Spence, personal communication) (*Takei et al., 2002*).

2.4. Kinetic Measurements of Protein Folding

2.4.1. Two-State Folding

A folding reaction can be analyzed by fitting folding and unfolding traces collected at various denaturant concentrations to exponential equations to obtain the observed rate of folding or unfolding at the respective denaturant concentrations (*Fersht, 1998*). When a folding reaction encounters only a single rate-limiting step, a single exponential equation is sufficient for fitting a folding or unfolding trace. When, as in Im9 and Im7, the folding of a small fraction of molecules is limited by peptide bond isomerization a double exponential function needs to be used to achieve a good fit (equation 1) (*Ferguson et al., 1999*).

$$F(t) = A_1 e^{-k_{obs1}t} + A_2 e^{-k_{obs2}t} + F_E \qquad (1)$$

F_E is the equilibrium fluorescence at the respective urea concentration, k_{obs1} and k_{obs2} are the observable rate constants and A_1 and A_2 the respective amplitudes. If no signal is lost during the dead time the fluorescence at time zero is expected to be $F_E+A_1+A_2$. When the fluorescence increases during the reaction, A (amplitude) takes on a negative value. Higher order exponential equations are necessary when several rate-limiting steps are encountered on the way from the unfolded to the native conformation. Barrier-less folding, without a high-energy transition-state, generates folding traces that do not fit to an exponential equation (*Huang et al., 2002; Onuchic et al., 1996; Sabelko et al., 1999*).

The depiction of the folding and unfolding rates as a function of denaturant concentration is called Chevron-plot or V-curve. When the logarithms of the folding and unfolding rates show a linear dependence on the denaturant concentration the resulting plot can be fitted to a two state transition (Figure 1.1) (*Fersht, 1998*).

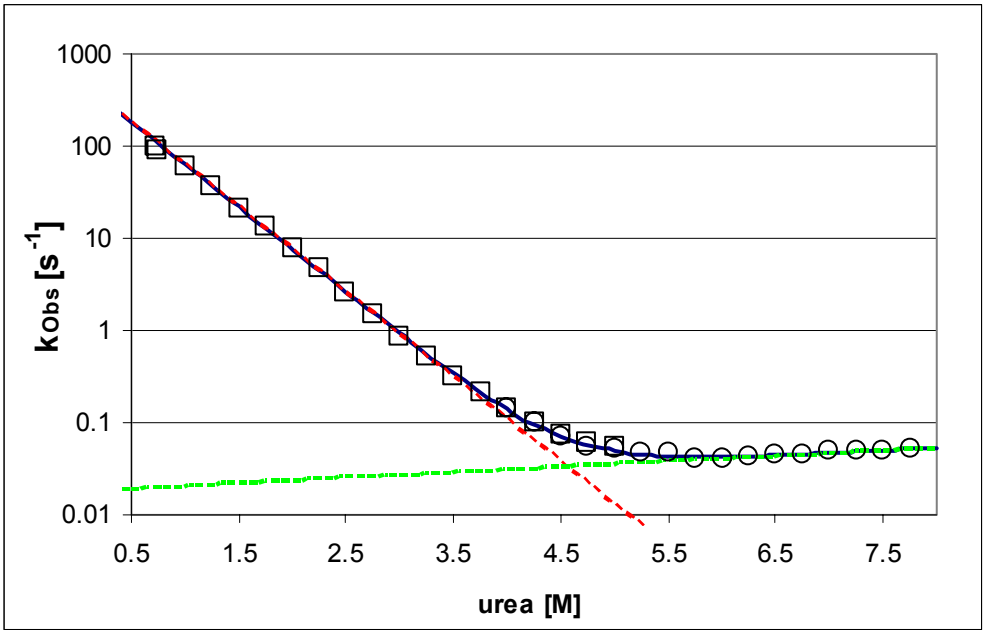


Figure 1.1 Chevron plot with linear refolding and unfolding branch. The squares are refolding rates and the circles are unfolding rates. The black line is a simulation of k_{Obs} (equation 4) with parameters that result in a curve that describes the measured data. The red hatched line represents the dependence of the refolding rate on denaturant concentration (equation 2) and the green hatched line represents the dependence of the unfolding rate on denaturant concentration (equation 3). The red hatched line (k_{UN} [denaturant]) crosses the y-axis at 530 s^{-1}, which is the refolding rate in water (0 M urea) (equation 2), its slope is -5 kJ mol^{-1} M^{-1} which is the dependence of the relative stability (to the unfolded state) of the transition state on the denaturant concentration (equation 5). The green hatched line (k_{UN} [denaturant]) crosses the y-axis at 0.013 s^{-1}, which is the unfolding rate in water (0 urea) (equation 3), its slope is 0.32 kJ mol^{-1} M^{-1} which is the dependence of the relative (to the native state) stability of the transition state on the denaturant concentration (equation 6).

In the case of two-state folding the dependence of the folding rate on denaturant concentration is described by:

$$\ln k_{UN} = \ln k_{UN}^{H_2O} - \frac{m_{UTS}[urea]}{RT} \qquad (2)$$

The denaturant dependence of the unfolding rate on denaturant concentration is described by:

$$\ln k_{NU} = \ln k_{NU}^{H_2O} + \frac{m_{NTS}[urea]}{RT} \qquad (3)$$

The observed rate constant k_{Obs} is described as:

$$k_{Obs} = k_{UN} + k_{NU} = \ln\left(k_{UN} e^{\frac{m_{UTS}[urea]}{RT}} + k_{NU} e^{\frac{-m_{NTS}[urea]}{RT}} \right) \quad (4)$$

The parameters obtained from this fitting procedure are k_{UN}, m_{UN}, k_{NU}, m_{NU}.

k_{UN} and k_{NU}, commonly measured in s^{-1}, are the folding rates and unfolding rates, respectively, from which the relative energy of the transition and native states can be calculated according to:

$$\Delta G_{UTS} = RT \ln\left(\frac{k_{UN}}{k_{\max}} \right) \quad (5)$$

$$\Delta G_{NTS} = RT \ln\left(\frac{k_{NU}}{k_{\max}} \right) \quad (6)$$

$$\Delta G_{UN} = \Delta G_{NTS} - \Delta G_{UTS} = RT \ln\left(\frac{k_{UN}}{k_{NU}} \right) = RT \ln\left(K_{UN} \right) \quad (7)$$

The pre-exponential term k_{max} describes the maximal folding rate according to:

$$k_{ij} = k_{\max} e^{\Delta G_{ij}} \quad (8)$$

Thus k_{max} represents the rate predicted to occur in the absence of any energy barrier. We use $4.8 \times 10^8\ s^{-1}$. The m-values indicate the sensitivity of a certain rate (equations 2, 3) or the resulting relative stability (equation 5, 6, 7) to denaturant concentration. As the denaturing effect of urea is based on the solvation of hydrophobic groups, the urea dependence of a rate or a relative stability is commonly interpreted as the change in accessible hydrophobic surface area during the respective reaction step. The absolute sum of the m-values is the total m-value:

$$m_{total} = -m_{UN} + m_{NU} \quad (9)$$

The total m-value represents the difference in surface exposed hydrophobic area between the denatured and the native state and gives an indication of the compaction of the native state relative to the denatured state. The compactness of the transition state relative to the native state can be calculated according to:

$$\beta_{TS} = \frac{-m_{UN}}{m_{total}} \quad (10)$$

Analysis of the folding amplitudes can provide additional fitting constraints. The folding amplitudes should be described by the same parameters as the rates from the respective experiments (Figure 1.2).

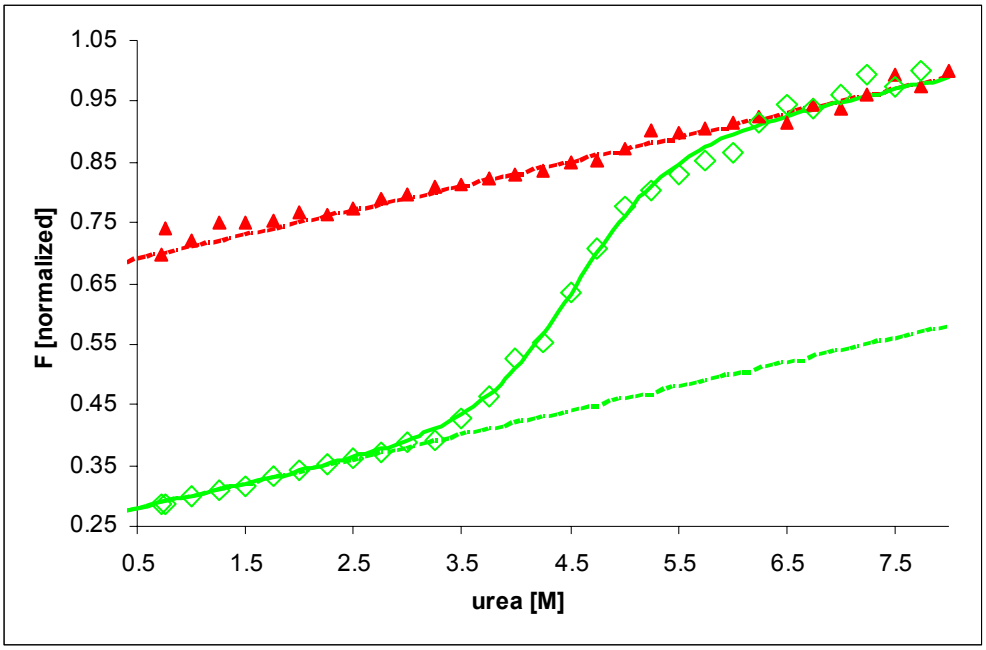


Figure 1.2 The equilibrium fluorescence (green diamonds) and the fluorescence of the unfolded state as a function of denaturant concentration (red triangles). The hatched lines represent the fluorescence of the native state (green hatched line) and unfolded state (red hatched line) and their relative dependence on urea concentration. The continuous line represents the equilibrium fluorescence calculated using equation (13) with F_N^{H20}=0.25, F_U^{H20} =0.67 as the fluorescence in water of the native and the unfolded state, respectively, and F_{Ndep}=0.04 M^{-1}, F_{Udep} =0.04 M^{-1} as the dependence on urea of the fluorescence of the native and unfolded state, respectively and the parameters obtained from the fitting of the rates (Figure 1). The fluorescence is normalized to the unfolded state in 8 M urea.

Additional parameters for describing the folding amplitudes are F_N^{H20} and F_U^{H20}, the fluorescence of the native state and denatured state in 0 M urea, respectively, and their relative dependence on urea concentration $F_N$dep and $F_U$dep, respectively (equation 11 and 12).

$$F_N = F_N^{H_2O} + F_{N\,dep}[urea] \tag{11}$$

$$F_U = F_U^{H_2O} + F_{U\,dep}[urea] \tag{12}$$

The equilibrium fluorescence F_E at each denaturant concentration is defined by the fluorescence of the native and the denatured state and the fraction of protein in the native state f_N and in the denatured state f_D at each concentration.

$$F_E = F_N f_N + F_U f_U \tag{13}$$

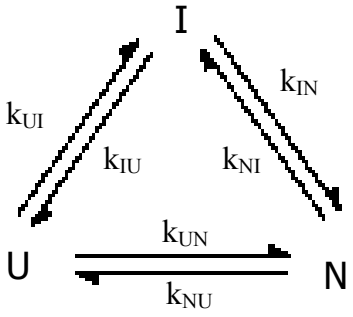
The fraction of protein in the native and denatured state can be calculated according to equations 14 and 15 using the parameters obtained from fitting the rates.

$$f_N = \frac{\frac{k_{UN}}{k_{NU}}}{1+\frac{k_{UN}}{k_{NU}}} = \frac{K_{UN}}{1+K_{UN}} \tag{14}$$

$$f_U = 1 - f_N \tag{15}$$

2.4.2. Three-State Folding

It occurs that the folding rates show a non-linear dependence on the denaturant concentration, which leads to a curved folding branch. This kind of Chevron plot or V-curve can be described by a simple two-state transition only if denaturant activity is assumed to be non-linear. Other mechanisms that can describe a curved folding branch are a switch in the rate-limiting transition state as the urea concentration increases, or the transient population of an intermediate during the dead-time of the instrument at low denaturant concentrations (*Sanchez and Kiefhaber, 2003*). Folding through an intermediate can be illustrated by a triangular mechanism (Scheme1). This scheme can be linearized by setting k_{IN} and k_{NI} to 0 (off-pathway intermediate), setting k_{UN} and k_{NU} to 0 (obligatory on-pathway intermediate), or setting k_{UI} and k_{IU} to 0 (over-folded intermediate) (*Baldwin, 1996*):



Scheme 1

As in the mechanism described previously each microscopic rate-constant, k_{ij} depends linearly on the denaturant concentration according to its m-value, m_{ij} (equations 2, 3). The observed rates can be described by three coupled equations (16, 17, and 18) (*Ikai and Tanford, 1973*):

$$\frac{d}{dt}[U] = -k_{UN}[U] - k_{UI}[U] + k_{NU}[N] + k_{IU}[I] \tag{16}$$

$$\frac{d}{dt}[N] = -k_{NU}[N] - k_{NI}[N] + k_{IN}[I] + k_{UN}[U] \tag{17}$$

$$\frac{d}{dt}[I] = -k_{IN}[I] - k_{IU}[I] + k_{NI}[N] + k_{UI}[I] \tag{18}$$

This system of equations can be written more conveniently as a rate-matrix, $\overline{\overline{A}}$ multiplied with a vector as seen in equation (19):

$$\frac{d}{dt}\begin{pmatrix} U \\ I \\ N \end{pmatrix} = \underbrace{\begin{bmatrix} -(k_{UN}+k_{UI}) & k_{IU} & k_{NU} \\ k_{UI} & -(k_{IN}+k_{IU}) & k_{NI} \\ k_{UN} & k_{IN} & -(k_{NI}+k_{NU}) \end{bmatrix}}_{\overline{\overline{A}}} \times \begin{pmatrix} U \\ I \\ N \end{pmatrix} \tag{19}$$

Introducing a diagonal matrix $\overline{\overline{\lambda}}$ that is equivalent to $\overline{\overline{A}}$ according to:

$$\overline{\overline{A}} \times \overline{y} = \overline{\overline{\lambda}} \times \overline{y} \tag{20}$$

$$\left(\overline{\overline{A}} \times \overline{y}\right) - \left(\overline{\overline{\lambda}} \times \overline{y}\right) = 0 \tag{21}$$

$$\left(\overline{\overline{A}} - \lambda\right)\left(\overline{y}\right) = 0 \tag{22}$$

and deriving the determinant of the resulting matrix which has to be equivalent to 0 according to:

$$\begin{bmatrix} -(k_{UN}+k_{UI}+\lambda) & k_{IU} & k_{NU} \\ k_{UI} & -(k_{IN}+k_{IU}+\lambda) & k_{NI} \\ k_{UN} & k_{IN} & -(k_{NI}+k_{NU}+\lambda) \end{bmatrix} \times \begin{pmatrix} U \\ I \\ N \end{pmatrix} = 0 \tag{23}$$

results in a quadratic equation (24) of the following type determining λ_1 and λ_2, the two rate-constants observable in the triangular mechanism 1.

$$f(\lambda) = a\lambda^2 + b_1\lambda + c = 0 \tag{24}$$

The solution of this equation is:

$$\lambda_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{25}$$

with:

$$a = 1$$

$$b = \sum_{j \neq i} k_{ij}$$

$$c = g_1 + g_2 + g_3$$

$$g_1 = \left(k_{NI} k_{IU}\right) + \left(k_{NU} k_{IU}\right) + \left(k_{IN} k_{NU}\right)$$

$$g_2 = \left(k_{NU} k_{UI}\right) + \left(k_{UI} k_{NI}\right) + \left(k_{UN} k_{NI}\right)$$

$$g_3 = \left(k_{UI} k_{IN}\right) + \left(k_{IU} k_{UN}\right) + \left(k_{UN} k_{IN}\right)$$

Figure 1.3 shows a fit according to the scheme described above. The symbols are the observable rate-constants measured in a conventional stopped-flow apparatus. The red and the blue lines represent the microscopic rate constants at each urea concentration. The nonlinear function that overlays with the data is the observable rate-constants which result from the depicted microscopic rate-constants and their dependence on urea.

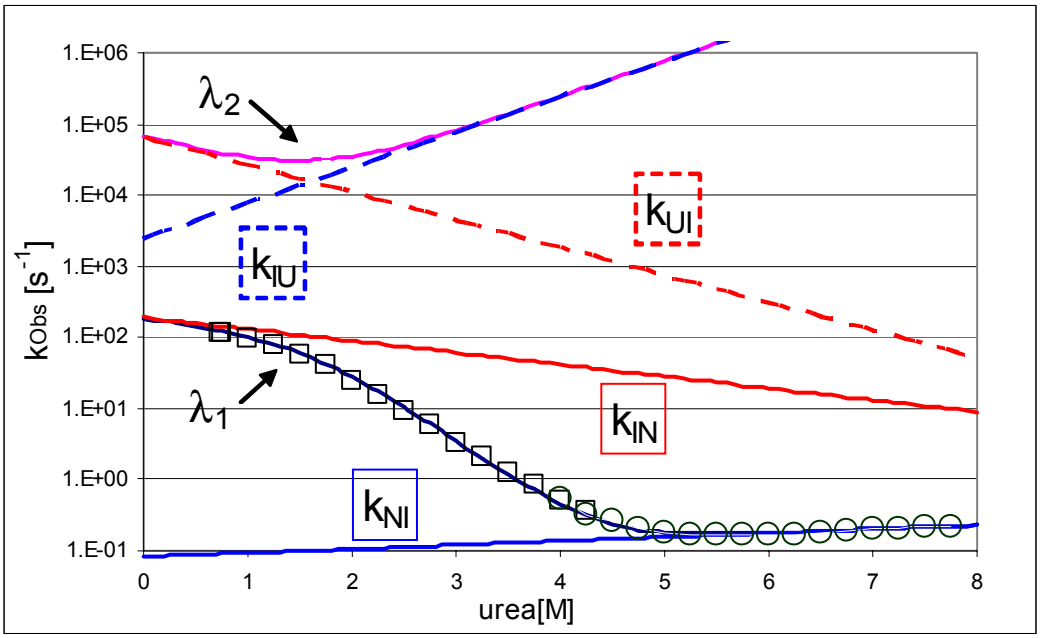


Figure 1.3 Fitting a curved refolding branch to a model involving a three-state on-pathway intermediate. The squares are folding rates and the circles are unfolding rates. The black continuous line is a simulation of λ_1 (equation 25) with parameters that result in a curve that describes the measured data. The magenta continuous line is a simulation of λ_2 (equation 25) for which no experimental data is shown. The red hatched line represents the dependence of k_{UI} on denaturant concentration; the blue hatched line represents the dependence of k_{IU} on the denaturant concentration. The red continuous line represents the dependence of k_{IN} on denaturant concentration; the blue continuous line represents the dependence of k_{NI} on the denaturant concentration. In accordance with thermodynamic principles the data presented here can be fit to an on-pathway model as well as to an off-pathway model and only the equilibrium stability of the intermediate and its equilibrium m-value is known. Here the on-pathway fit is shown. The stability of the intermediate is 7.7 kJ mol^{-1} and its equilibrium m-value is 4.8 kJ $mol^{-1}M^{-1}$. The rates for the folding and unfolding of the native state from or to the intermediate and their urea-dependence can be directly determined from the shown data, k_{IN}= 190 s^{-1}, m_{IN}= -0.9 kJ $mol^{-1}M^{-1}$, k_{NI}= 0.082 s^{-1}, m_{NI}=0.3 kJ $mol^{-1}M^{-1}$. The stability of the native state is 26.0 kJ mol^{-1}. The relative compaction of the intermediate state and the second transition state is β_I= 0.8 and β_{TS2}=0.95.

From the microscopic rate constants and the equilibrium constants obtained in this analysis the stability of the intermediate, the second transition-state and the native state can be determined according to:

$$\Delta G_{UI} = RT \ln(K_{UI}) \tag{26}$$

$$\Delta G_{UN} = RT \ln(K_{UI} K_{IN}) \tag{27}$$

$$\Delta G_{ITS} = RT \ln \frac{k_{ITS}}{4.8\times10^{8}} \tag{28}$$

$$\Delta G_{UTS} = \Delta G_{ITS} + \Delta G_{UI} \tag{29}$$

The β-values as indicators of the compaction of the respective state relative to the native state can be calculated according to:

$$\beta_{TS} = \frac{m_{UI}^{equilibrium}}{m_{total}} \tag{30}$$

$$\beta_{I} = \frac{m_{UI}^{equilibrium} + m_{IN}}{m_{total}} \tag{31}$$

As in the analysis according to a two-state transition the folding amplitudes can provide additional constraints for fitting, especially when the fluorescence of the intermediate is different from the fluorescence of the unfolded state. In addition to equations 11 and 12, equation 32 is introduced, which describes the fluorescence of the intermediate state in 0 M urea, $F^{H20}{}_I$ and its relative dependence on urea concentration, $F_I$dep.

$$F_I = F_I^{H_2O} + F_{I\,dep}\left[urea\right] \tag{32}$$

The fluorescence at equilibrium at each urea concentration is defined by:

$$F_E = F_N f_N + F_I f_I + F_U f_U \tag{33}$$

$F_{(N,I,U)}$ is the fluorescence of each state at the respective urea concentration and *f* the fraction of molecules in the respective state at equilibrium. The fraction of native, intermediate and denatured protein at equilibrium can be defined by the rate constants at the respective urea concentration according to:

$$f_N = \frac{k_{UI}k_{IN}}{\lambda_1\lambda_2} \tag{34}$$

$$f_I = \frac{k_{UI}k_{NI}}{\lambda_1\lambda_2} \tag{35}$$

$$f_U = \frac{k_{IU}k_{NI}}{\lambda_1\lambda_2} \tag{36}$$

The initial fluorescence (F_E+A_1+A_2) is here not the fluorescence of the denatured state under equilibrium conditions, but the fluorescence of the denatured, intermediate and native states multiplied with the fraction of molecules in the respective conformation (equation 37) after elapse of the dead-time added to the equilibrium fluorescence F_E at the respective urea

concentration. The fluorescence after the dead-time, t, i.e. after the reaction with the observable rate constant λ_2 has come to completion is thus defined by:

$$F_{Initial} = F_N\left(f_{N(t)} + f_N\right) + F_U\left(f_{U(t)} + f_U\right) + F_I(f_{I(t)} + f_I) \tag{37}$$

Where equations (38-40) describe which fraction of the molecules are in the native, intermediate and denatured state, respectively.

$$f_{N(t)} = \frac{k_{UI}k_{IN}}{\lambda_2\left(\lambda_1 - \lambda_2\right)}e^{-\lambda_2 t} \tag{38}$$

$$f_{I(t)} = \frac{\lambda_2 k_{UI} - k_{UI}k_{NI}}{\lambda_2\left(\lambda_1 - \lambda_2\right)}e^{-\lambda_2 t} \tag{39}$$

$$f_{U(t)} = \frac{-\lambda_2 k_{UI} + k_{UI}k_{NI} + k_{UI}k_{IN}}{\lambda_2\left(\lambda_1 - \lambda_2\right)}e^{-\lambda_2 t} \tag{40}$$

The folding amplitudes that were obtained in the experiment that provided the rates shown in the Figure 1.3 are shown in Figure 1.4. The fit that is shown is obtained using the rates and m-values obtained from the rates and the fluorescence of each state.

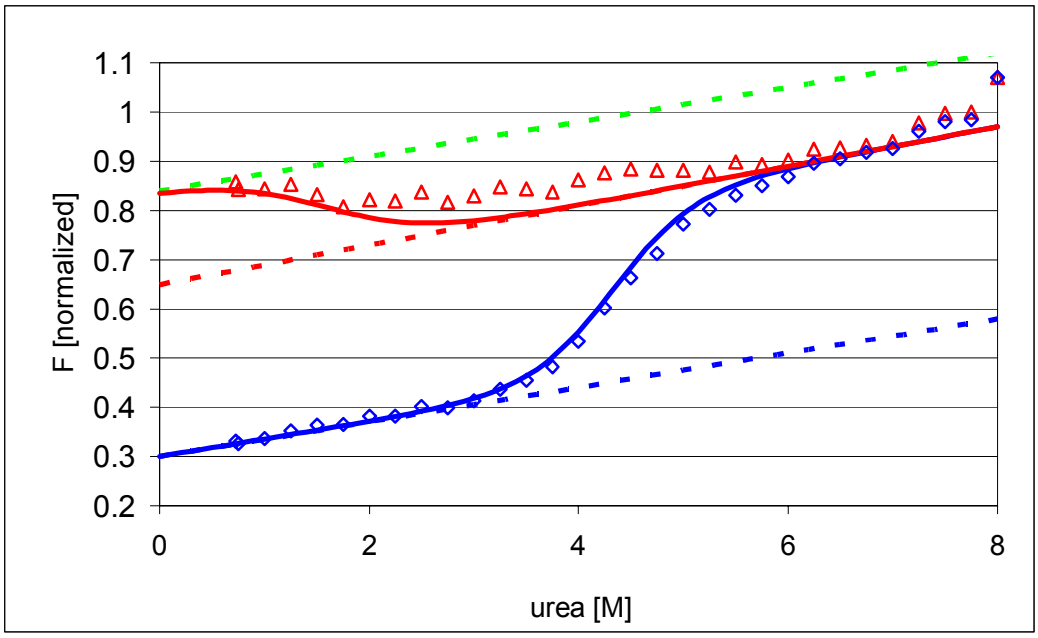


Figure 1.4 The equilibrium fluorescence (blue diamonds) and the initial fluorescence after the dead-time (red triangles). The hatched lines represent the fluorescence of the native state (blue hatched line), intermediate state (green hatched line) and denatured state (red hatched line) and their relative dependence on urea concentration. The continuous blue line represents the equilibrium fluorescence calculated using equation (33). The continuous red lien represents the initial fluorescence calculated using equation 37-40. The calculated fluorescence values were obtained using F_N =0.3, F_I=0.84, F_U =0.65, F_{Ndep}=0.035 M^{-1}, F_{Idep} [urea] =0.035 M^{-1} and F_{Udep} [urea] =0.04 M^{-1} and the parameters obtained from the fitting of the rates (Figure 1.3).

2.5. The Protein Engineering Method

Measuring the folding and unfolding rates of proteins under different conditions at varying denaturant concentrations allows us to calculate the relative energy of various states and their relative compaction. How can structural information about these states, some of them transient, be derived? The protein engineering method that was largely developed by Fersht et al. makes an indirect structural characterization of the rate-limiting state and of transient intermediates possible. Conservative mutations that usually involve truncation of a hydrophobic side-chain are made at selected positions throughout the molecule. A Chevron-plot for each protein including the wild-type is measured and thus the stability the different states in the mutants and the wild-type protein can be determined and compared. From this information conclusions about the role of each residue in the various states are drawn and quantified by the so-called Φ-value (*Fersht et al., 1992*). Qualitatively speaking, truncating a residue that does not stabilize

the transition state but does stabilizes the native state, should globally destabilize the native state but not the transition state (Φ-value=0). Truncating a residue that stabilizes the transition state just as much as the native state should globally stabilize both of these states by the same amount (Φ-value=1) (Figure 1.5). Structurally a Φ-value of 1 is interpreted as the truncated residue forming all native contacts in the transition state and a Φ-value of 0 is taken to indicate that the truncated residue makes no stabilizing interactions in the transition state. Φ-values between 0 and 1 are considered normal Φ-values. Abnormal Φ-values are negative values and those over 1. In the former case the truncated side-chain has a different role in the transitions state than it has in the native state, stabilizing for example the native state but destabilizing the transition state. The latter case is taken to indicate that a certain residue forms more stabilizing interactions in the transitions state than in the native state. Logically some of these interactions must be non-native.

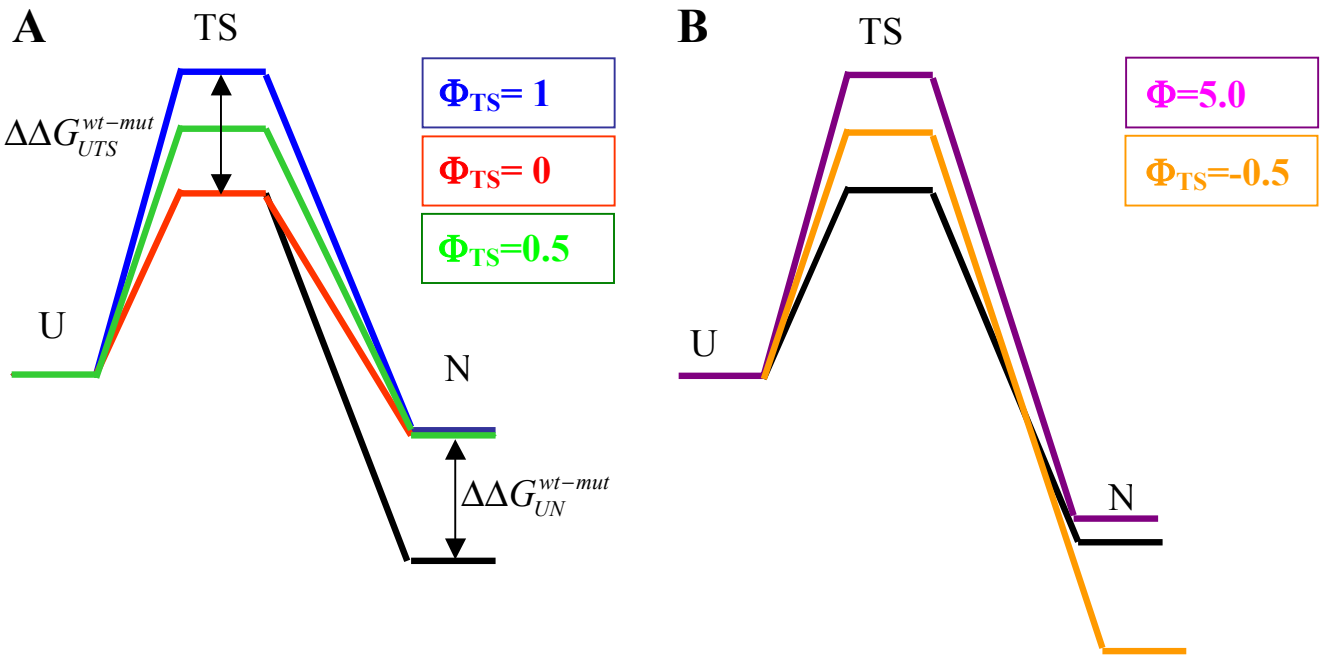


Figure 1.5 Schematic energy diagrams illustrating normal (A) and abnormal (B) Φ-values. The energy diagram of the wild-type is shown in black and the various mutants shown in color. The Φ-values are shown in the color of the respective energy diagram.

Quantitatively the Φ-values express the energy change recorded for the transition state divided by the energy change recorded for the native state. In a two state system Φ_{TS} can be calculated by equation 41.

$$\Phi_{TS} = \frac{\Delta\Delta G_{UTS}^{wt-mut}}{\Delta\Delta G_{UN}^{wt-mut}} = \frac{\Delta G_{UTS}^{wt} - \Delta G_{UTS}^{mut}}{\Delta G_{UN}^{wt} - \Delta G_{UN}^{mut}} \quad (41)$$

A similar analysis is possible when an intermediate is observed. The Φ_I-values are then calculated by dividing the energy changes recorded for the intermediate state by the energy changes recorded for the native state:

$$\Phi_{I} = \frac{\Delta\Delta G_{UI}^{wt-mut}}{\Delta\Delta G_{UN}^{wt-mut}} = \frac{\Delta G_{UI}^{wt} - \Delta G_{UI}^{mut}}{\Delta G_{UN}^{wt} - \Delta G_{UN}^{mut}} \quad (42)$$

In this case two Φ_{TS}-values may be obtained, one for each transition-state.

Truncating hydrophobic residues and measuring the destabilization of the states in which these hydrophobic residues make stabilizing contact is one way of using protein engineering to structurally characterize unstable states. Secondary structure formation, which is characterized by local hydrogen bonds without clear consolidation of non-local hydrophobic contacts, cannot be probed in this manner. An ingenious way to measure the consolidation of secondary structure in a transient state is Ala to Gly mutations at solvent exposed positions, employing either naturally occurring alanines or newly introduced alanines (*Burton et al., 1997; Burton et al., 1998*). This strategy is based on the high entropic cost of fixing the Φ- and Ψ–angles of glycine to values that are appropriate for a helix. This high entropic cost is the consequence of the very large conformational space that is open to glycine in the unfolded state. The conformational freedom of alanine in the unfolded state is much smaller; therefore the entropic costs for fixing the Φ- and Ψ–angles to helical values is also much smaller. Introducing a glycine for an alanine entropically stabilizes all conformations in which the affected residues is not involved in a helix and causes a relative destabilization of all conformations in which this residue is involved in an element of secondary structure requiring fixed Φ- and Ψ–angles.

2.6. Model Systems for Protein Folding

The presented work mainly treats the mechanism by which the unique native structure is reached and should be understood as an investigation into protein folding kinetics. During the folding reaction a protein transiently populates one or several extremely unstable states. The observation of such states in itself is a challenge, their structural and thermodynamic characterization and the investigation of the mechanism by which such states become

populated and depopulated even more so. There are certain criteria that make the folding reaction of a sequence amenable to structural, thermodynamic and kinetic investigations.

Firstly, the folding and unfolding reactions should be fully reversible and reproducible in-vitro, making the reactions accessible to physico-chemical characterization (*Pain, 1994*).

Secondly, analysis of the observed reactions is facilitated if the observed transitions involve only non-covalent interactions rather than e.g. the formation and reduction of disulfide bonds between topologically distant cysteine residues.

Thirdly, it is preferable that the protein should not contain proline residues, which can give rise to complicated folding kinetics. In a polypeptide, the amino acids are covalently linked by amide bonds, which can occur in a cis as well as a trans-conformation. Because of steric hindrances the amide bonds in unfolded as well as in folded proteins usually occupy the trans-conformation. The only exception to this rule is proline. Because the δ-carbon atom of the prolyl sidechain forms a covalent bond to the amide nitrogen the geometry of the peptide bond that is amino terminal to a proline is very different from the other peptide bonds. The cis/trans-isomerization state of this peptide bond in the unfolded state depends on the local sequence; typically each proline peptide bond adopts the trans-conformation with a likelihood of 60%. However, in the native state there is a unique native isomerization state for each proline residue. In a folding reaction with one or several prolines the population of molecules that happen to have all peptide bonds in the native isomerization state fold quickly, whereas the folding of polypeptide chains that contain one or several incorrect prolyl isomers is limited by the isomerization of these peptide bonds and gives rise to one or several slow phases (*Adler and Scheraga, 1990; Brandts et al., 1975*).

Fourthly, the folding and unfolding reactions as they are triggered by changes in buffer conditions or temperature have to occur on a timescale that is resolvable by current methods. The dead time of conventional stopped-flow methods is approximately 2-4 ms; continuous flow methods can achieve a higher resolution with dead-times as low as 10μs. The highest time-resolution to date can be achieved using pH and temperature jump methods (*Eaton et al., 1996*). These methods are limited to proteins where folding and/or unfolding reactions can be triggered by temperature jumps of technically feasible size.

Fifthly, the conformational changes that occur in folding and unfolding reactions should give rise to a signal that can be monitored. A tryptophan that is buried in the native state increases its intrinsic fluorescence as tertiary interactions are consolidated and it is transferred to the hydrophobic core of the protein. The fluorescence increase during folding of a preferably unique buried tryptophan can be monitored with relatively simple equipment.

2.7. Immunity Proteins as a Biological System

In order to be competitive in nutrient-limited environments bacteria express a variety of bactericidal molecules. *Escherichia coli* express colicins as part of the SOS reaction. Although colicins are specific for *E. coli,* similar toxins are found in other gram-negative bacteria. The nine known E-group colicins are made up of three domains, one N-terminal domain that aids translocation into the target cell, a central domain that binds the colicin to the membrane of the target cell by interacting with its vitamin B_{12} receptor, and the C-terminal enzymatic cytotoxic domain. Based on their mechanism of cytotoxicity E-group colicins can be divided into different subgroups, pore-forming colicins, endonucleases (DNases) or ribonucleases (RNases). A host bacterium expressing a colicin needs to protect itself against the cytotoxic activity of the enzymatic domain. This is accomplished by co-expression of a specific colicin inhibitor. Each of the known E-group colicins (E1-E9) has its cognate inhibitor (Im1-Im9). Studies on E9 and Im9 indicate that the inhibitor folds to its native structure in less than 1 ms and binds to the colicin at a rate close to the diffusion limit and with a K_D of 10^{-16} M. Although only the structures of the E3-Im3, E7-Im7 and E9-Im9 complexes have been solved, sequence alignments and mutagenic studies of the activity of colicins and its inhibition by immunity proteins reveal a striking similarity between the mechanisms of inhibition. Immunity proteins do not bind to the active site of the colicins, but rather to sequences N-terminally adjacent to the active site. Inhibition is the result of steric and/or electrostatic repulsion between the negatively charged immunity protein and the substrate of the colicin, DNA or RNA, which is also negatively charged and moreover a very bulky polymer. This mechanism of action has at least two biological advantages: Firstly the immunity proteins can bind to the colicins co-translationally even before the active site becomes translated; secondly a bacterium has the possibility to mutate the binding site for the immunity protein while the active site, responsible for the cytotoxicity, remains unchanged. Through co-evolution between colicin and immunity protein the host bacterium can generate novel colicins and immunity proteins. The chance that the target bacteria already have a cognate immunity protein for such new colicins is small.

In the case of the well-studied DNase colicins and their immunity proteins a very small number of amino acids control the specificity of the immunity protein and these residues make only minor contributions to the binding energy. The highly conserved residues that are responsible for the high affinity are separate from the variable residues that confer specificity. This has been interpreted in terms of a "dual recognition mechanism" in which some highly conserved residues provide most of the favorable binding energy and the variable specificity residues modulate this binding energy by making neutral, favorable on unfavorable contributions (*Kleanthous and Walker, 2001*).

2.8. Immunity Proteins as a Model System for Protein Folding

The principle question in protein folding is how the sequence and/or topology of a protein determine the structural and thermodynamic characteristics of the native state and the mechanism by which this state is reached. One way of addressing these questions is a homology study. For this purpose two members of the family of immunity proteins, Im7 and Im9 were characterized in great detail (*Capaldi et al., 2002; Capaldi et al., 2001; Ferguson et al., 1999; Ferguson et al., 2001; Friel et al., 2003; Gorski et al., 2001 Paci 2004*). These two proteins show only approximately 60% sequence identity but almost identical topology. Essentially, they can be described as distorted four-helix bundles (Figure 1.6). The first two helices (in Im9 12-23 and 30-44, respectively) pack against each other forming an antiparallel two-helix structure joined by a short six-residue turn. The third helix is extremely short (in Im9 50-55) but forms the topological center of the structure. Helix 3 interacts with the two-helix bundle formed by helices 1 and 2 and provides a connection between this structure and helix 4. Helix 4 (65-77) interacts predominantly with helix 3 and with the N-terminal region of helix 1 (Figure 1.6).

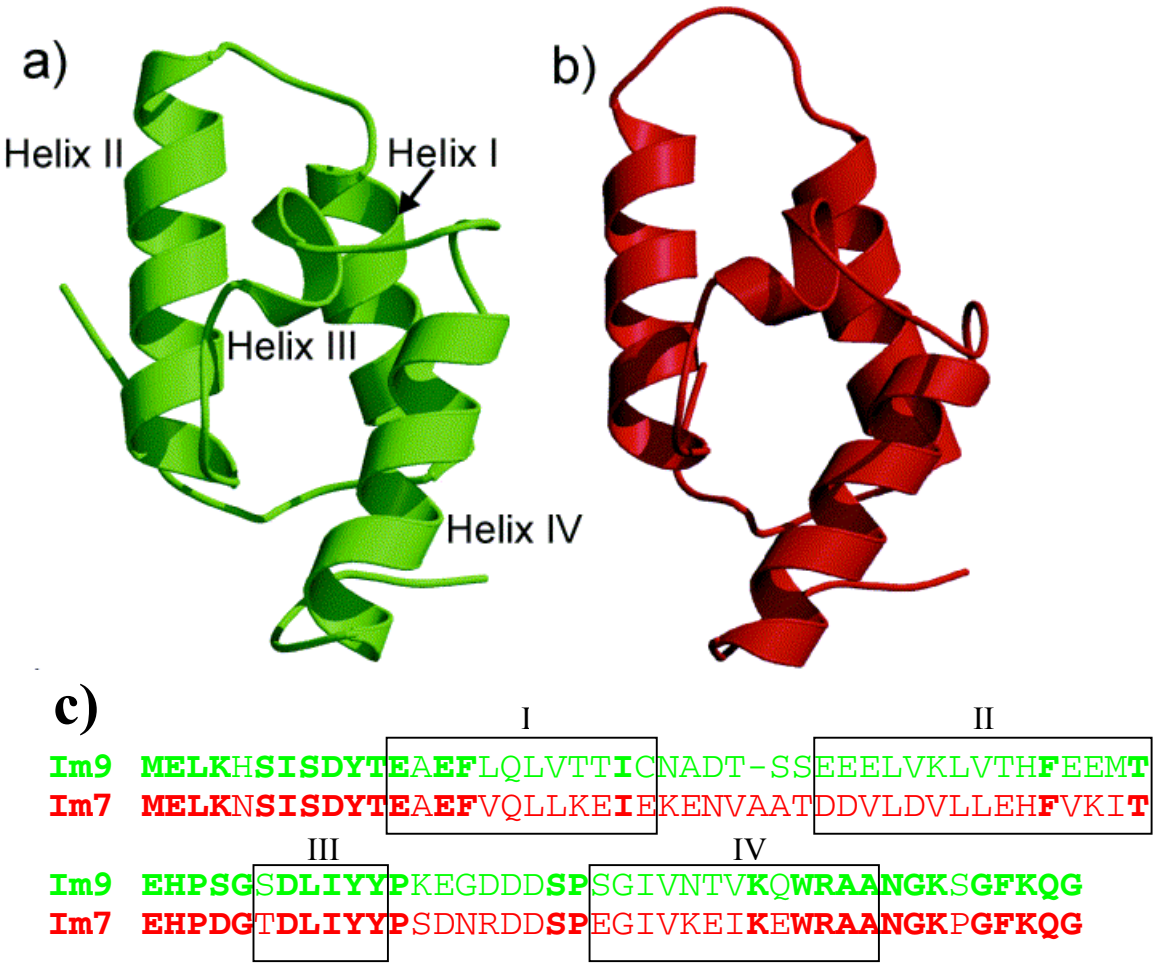


Figure 1.6 Comparison of the structures and sequences of Im9 and Im7. a) Im9 and b) Im7 in the same orientation. c) Sequence alignment of Im9 (red) and Im7 (green) where identical residues are shown in bold. The position of the helix 1 (12-23), helix 2 (30-44), helix 3 (50-55) and helix 4 (65-77) are indicated by black boxes. Figure modified from Friel et al 2003.

Keeping the above criteria in mind we can evaluate whether immunity proteins are a suitable model system to study folding. Immunity proteins are quite small (85-87 residues) and easier to handle than many large multidomain proteins. They unfold reversibly under the influence of chemical denaturants such as urea and guanidinium chloride. Immunity proteins lack disulfide bonds, cis-proline residues in the native state and prosthetic groups (*Ferguson et al., 1999*); the development and consolidation of the native structure which is based solely on intramolecular non-covalent interactions and the hydrophobic effect can be studied in the absence of complicating factors. The folding reaction of the immunity proteins can be monitored by two different spectroscopic methods. Firstly, the proteins' α-helical topologically makes them

amenable to equilibrium and kinetic studies by far-UV circular dichroism. Secondly, there is a unique and conserved tryptophan (W74 in Im9) the fluorescence of which is slighlty quenched by the buffer when the protein is unfolded. In the folded state the tryptophan is very close to a histidine in the loop between helix 2 and helix 3 (H46 in Im9), and its fluorescence becomes highly quenched. The fluorescence decrease of this tryptophan is a good probe of global folding and unfolding. The equilibrium stability of the folded state and its denaturant sensitivity were measured by monitoring either the negative far UV-CD signal at 225 nm or the fluorescence of the tryptophan. Measurements of the stability and denaturant sensitivity of the native state by either spectroscopic probe indicate that the decrease of tryptophan fluorescence and the decrease of the far-UV CD signal at 225 nm monitor the same global unfolding event (*Ferguson et al., 1999; Gorski et al., 2001*). Thermodynamic measurements at 50 mM Na_2HPO_4 (pH 7.0), 2 mM DTT, 10° C, using urea as a denaturant show that Im9 and Im7 have similar equilibrium m_{un}-values of 4.62 kJ $M^{-1}mol^{-1}$ and 5.21 kJ $M^{-1}mol^{-1}$, respectively. Under these conditions Im7 (16.8 kJ mol^{-1}) is much less stable than Im9 (26.1 kJ mol^{-1}) (*Ferguson et al., 1999*). The kinetic traces of folding monitored by stopped-flow fluorescence show that the folding of both proteins is well described by a double exponential equation. The slower phase has a rate-constant of 7×10^{-3} s^{-1} for both proteins and is independent of the final urea concentration of the folding buffer (*Ferguson et al., 1999*). The rate of the slow folding phase, its relative urea independence and its small amplitude suggest that some molecules refold from conformations in which one or both proline residues are in the non-native cis-conformation and the folding is limited by the isomerization of these peptide bonds. The unfolding data that can be recorded only at higher concentrations of denaturants fit well to a single exponential equation.

2.8.1. Two-State Folding of Im9

The folding and unfolding rates of Im9 show a linear dependence on urea concentration resulting in a Chevron plot with linear folding and unfolding branches. This result can be analyzed in terms of a cooperative two-state transition. The equilibrium stability and equilibrium m-value deduced from the kinetic measurements using tryptophan fluorescence as a spectroscopic probe compare well with the parameters from equilibrium denaturation curves that were recorded using the far UV-CD signal of the α-helices (*Ferguson et al., 1999; Gorski et al., 2001*). Analysis of the folding amplitudes shows that the protein refolds from an unfolded state with the expected spectroscopic properties and that the fluorescence at

equilibrium at each urea concentration is described by a two-state transition using the very parameters that were used for fitting the observed folding and unfolding rates (Figure 1.7). These observations make a very strong case for Im9 folding in a two-state manner. From kinetic measurements the position of the transition state on the reaction coordinate can be determined in form of the $\beta^{\ddagger}$-value. For Im9 the $\beta^{\ddagger}$ value is 0.94 showing that the transition state buries 94% of the hydrophobic surface that becomes buried in the native state and is therefore highly compact (*Ferguson et al., 1999*).

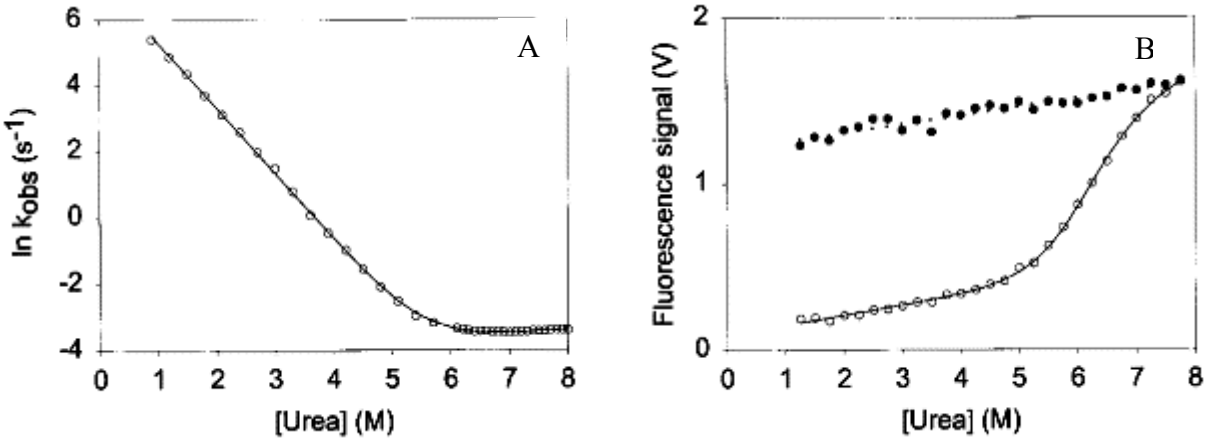


Figure 1.7 Denaturant dependence of the folding and unfolding of Im9. A) The natural logarithm of the rate constants is plotted against the denaturant concentration at 50 mM Na_2HPO_4 (pH 7.0, 2 mM DTT, 10 ˚C. B) The refolding amplitudes are plotted against the denaturant concentration. The continuous line is the best fit of the refolding rates and the refolding amplitudes to a two-state model (see Chapter: Analysis of kinetic measurements). The empty symbols represent the fluorescence after the folding reaction came to completion, F_E. The filled symbols represent the sum of F_E and the amplitudes obtained at the respective urea concentration, $F_{initial}$. Parameters used for fitting are: k_{UN}= 1450 s^{-1}, m_{UN}= -4.68 kJ $mol^{-1}M^{-1}$, k_{NU}= 0.0124 s^{-1}, m_{NU}= 0.3 kJ $mol^{-1}M^{-1}$.

The protein engineering method provides a technique for characterizing the interactions that stabilize the transition state and suggests a possible structure for this high-energy state (*Friel et al., 2003*). The work of Friel et al. suggests that the transition-state of Im9 contains the three long helices and that their hydrophobic surfaces are buried correctly, while the residues that are surface exposed in the native state remain exposed also in the transition state. There is no indication that the residues of helix 3 are in any way involved in stabilizing the rate-limiting transition state. The residues with the highest Φ_{TS}-values are those involved in the interface between helices 1 and 4, namely residues in the respective amino-terminal regions of these helices. Φ_{TS}-values for mutations in helices 1 and 2 decrease toward the carboxy-terminal end

of the helices, in fact very low Φ-values for Ala to Gly mutations in the carboxy-terminus of helix 4 indicate that this helix may be frayed or incompletely formed in the transition-state. According to its position, very far away from the strong interactions between helices 1 and 4, Φ_{TS}-values are very small in helix 2. Generally speaking all of the measured Φ_{TS}-values are small and none of them is close to unity. This supports a rather broad transition-state ensemble, encompassing numerous compact conformations, a suggestion that was confirmed by Paci et al. who used molecular dynamics, the native state structure and the experimental Φ_{TS}-values to simulate the transition state ensemble (*Paci et al., 2004*).

2.8.2. Three-State Folding of Im7

Kinetic experiments on Im7 indicate that its folding behavior at 50 mM Na_2HPO_4 (pH 7.0), 2 mM DTT, 10˚ C cannot be characterized by a two-state transition (*Ferguson et al., 1999*). The folding rate of Im7 does not show a linear dependence on the urea concentration. At low denaturant concentration, the observed rates are smaller than the rates expected from linear extrapolation of the folding rates at higher urea concentrations, an observation commonly termed roll-over (Figure 1.8). Analysis of the folding amplitudes determines that the protein refolds from a state with far higher fluorescence than expected from the unfolded state, an observation described as burst-phase amplitude. Furthermore the parameters obtained from a two-state fit to the equilibrium denaturation data do not agree with the parameters that describe the Chevron plot of Im7 (*Gorski et al., 2001*). This behavior is commonly attributed to three-state folding where an intermediate that is more stable and encounters a higher energy barrier than any other state before it becomes transiently populated at low denaturant concentrations. The accumulation of the intermediate of Im7 in folding experiments at low denaturant concentrations is very fast and occurs within the dead-time of the stopped flow instrument. The burst phase amplitude is the result of the folding occurring from a hyperfluorescent intermediate rather than from the unfolded state. The smaller slope of folding rate versus denaturant concentration that causes the rollover also indicates that the folding reaction occurs from a state in which the solvent accessible hydrophobic surface is smaller than in the denatured state. The kinetic m-value of each step of the reaction indicates how much hydrophobic surface is buried in the respective step. When the first observable step is folding from a structured intermediate to the second transition state the loss in solvent-exposed hydrophobic surface is naturally smaller than when the second transitions state is reached from the unfolded state.

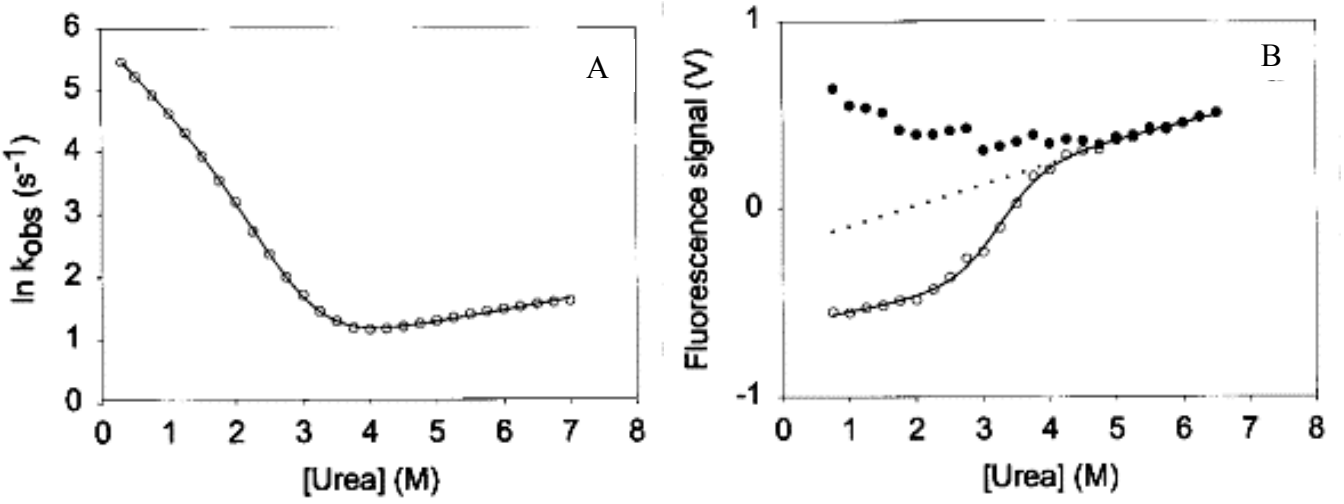


Figure 1.8 Denaturant dependence of the folding and unfolding of Im7. A) The natural logarithm of the rate constants is plotted against the denaturant concentration at 50 mM Na_2HPO_4 (pH 7.0), 2 mM DTT, 10 ˚C. B). The refolding amplitudes are plotted against the denaturant concentration. The continuous line is the best fit of the refolding rates and the refolding amplitudes to a two-state model (see Chapter: Analysis of kinetic measurements). The empty symbols represent the fluorescence after the folding reaction came to completion, F_E. The filled symbols represent the sum of F_E and the amplitudes obtained at the respective urea concentration. The dotted line is the denaturant dependence of the unfolded state according to equation (12). Parameters used for fitting are: K_{UI}= 18, m_{UN}= 3.45 kJ $mol^{-1}M^{-1}$, k_{IN}= 345 s^{-1}, m_{ITS2}= 2.42 kJ $mol^{-1}M^{-1}$, k_{NI}= 1.43 s^{-1}, m_{NTS2}= 0.43 kJ $mol^{-1}M^{-1}$.

It has been suggested that a roll-over and a burst-phase amplitude can also be caused by transient aggregation; however the folding behavior of Im7 is independent of protein concentration. A rollover in the folding branch can theoretically also be fitted to non-linear denaturant action or a switch in transition states, practically however only the latter is possible. Stringent experiments have shown that a simultaneous fit to the rates and the folding amplitudes at various pHs can only be achieved by a three-state model involving an obligatory intermediate (Gorski 2001). In the described stopped-flow experiments only the accumulation of the intermediate in the dead-time of the instrument is observed and only the equilibrium stability of the intermediate can be determined. The determined rates for the slower rate-constant λ_2 can be fit by an on-pathway as well as by an off-pathway model for three-state folding in accordance with thermodynamic principles (Figure 1.9). However if formation of an intermediate is followed directly using a continuous-flow mixing device with a dead-time of approximately 100 µs it is possible to distinguish an on-pathway from an off-pathway

intermediate. Thus, it could be shown that the intermediate populated during the folding reaction of Im7 is on-pathway (Figure 1.9) (Capaldi 2001).

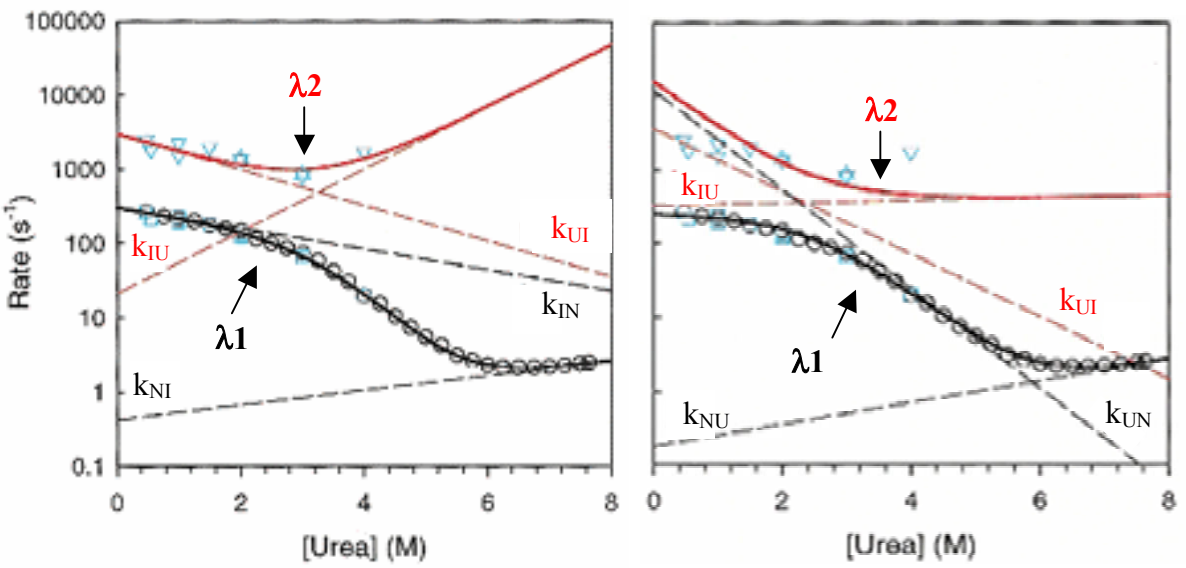


Figure 1.9 On- and off-pathway fits for the two rates observable in Im7 refolding experiments with a continuous flow mixer. The rates are shown as a function of urea concentration. The black symbols represent the data points that were collected using conventional stopped-flow equipment. The blue points are the rate constants determined from the merged stopped-flow and continuous-flow data. The slow phase is depicted by squares and the fast phase by triangles. Where replicate measurement didn't overlay very well they are depicted as upright triangles. A) The best fit of the data to an on-pathway folding model. The parameters obtained are k_{UI}=3000 s^{-1}, m_{UI}=1.3 kJ $mol^{-1}M^{-1}$, k_{IU}=21.6 s^{-1}, m_{IU}=-2.3 kJ $mol^{-1}M^{-1}$, k_{IN}=309 s^{-1}, m_{IN}=0.8 kJ $mol^{-1}M^{-1}$, k_{NI}=0.43 s^{-1}, m_{NI}= -0.5 kJ $mol^{-1}M^{-1}$, ΔG_{UN}= 27.1 kJ mol^{-1}, ΔG_{UI}=11.6 kJ mol^{-1}. B) The best fit of the data to an off-pathway model. The parameters obtained are k_{UI}=3600 s^{-1}, m_{UI}=2.3 kJ $mol^{-1}M^{-1}$, k_{IU}=330 s^{-1}, m_{IU}=-0.1 kJ $mol^{-1}M^{-1}$, k_{UN}=1200 s^{-1}, m_{UN}=3.7 kJ $mol^{-1}M^{-1}$, k_{NU}=0.18 s^{-1}, m_{NU}= -0.8 kJ $mol^{-1}M^{-1}$, ΔG_{UN}= 26.1 kJ mol^{-1}, ΔG_{UI}=5.63 kJ mol^{-1}.

As in Im9 the protein engineering method was employed to characterize the transient states that are not accessible to conventional spectroscopic methods. The structure of the second rate-liming transition state is very similar to the single rate-limiting state observed in Im9 folding. Interestingly, the Φ_{TS}-values are in general significantly higher than observed in the Φ-analysis of Im9 apparently contradicting the β–value of the transition-states, which indicate that the transition state of Im9 is slightly more compact. The analysis of the effect of typical Φ-mutations on the stability of the intermediate is especially interesting. At first glance the intermediate of Im7 appears structurally similar to the transition state of Im7 and Im9; helices

1, 2 and 4 are formed while residues in helix 3 do not make any stabilizing interactions. However, several Φ-values >1.0 are measured. Such abnormal Φ_I-values are observed for at least one mutation in each of the three long helices that are formed in the intermediate. Close inspection of all the Φ_I-values reveals that they are frequently larger than the equivalent Φ_{TS}-values suggesting that many of the hydrophobic interactions that stabilize the intermediate become disrupted in the transition state. Especially affected are residues at the interface of helix 1 and 2 and residues at the interface between helix 4 and the loop between helix 1 and 2. Truncating the hydrophobic side-chains that would be exposed in an entirely native intermediate destabilizes the intermediate. These observations make a strong case for a three-helix bundle intermediate that is in part stabilized by non-native interactions (*Capaldi et al., 2002*). An interpretation that is supported by the fact that a structure containing three helices docked in an entirely native way exposes a large patch of hydrophobicity to solvent. The tryptophan that shows only weak fluorescence when it is solvent-exposed in the unfolded state and is quenched by the histidine in the native state, is now buried in the hydrophobic core, but not quenched by histidine because the loop in which histidine lies is not yet structured.

An intermediate can also be observed in the folding reaction of Im9, depending on the buffer conditions. A generic way of stabilizing compact states is the addition of cosmotropic salts like Na_2SO_4. The intermediates of Im9 and Im7 can also be stabilized by lower pH. Conversely, increasing the pH from 7.0 to 8.0 in the presence of 0.4 M Na_2SO_4 destabilizes the intermediate observed in Im7 by 3.7 kJ mol^{-1}, a loss of stability of more than 30%. This suggests that formation of an intermediate is an intrinsic feature of the folding mechanism of immunity proteins. The buffer conditions merely determine whether this intermediate accumulates and becomes observable, meaning that it is more stable than the unfolded state and encounters a higher energy barrier than any other state before it. The intermediate of Im9 could be unobservable under standard conditions because either it is less stable than the denatured state or because it does not encounter as high an energy barrier as the denatured state encounters. The former could occur because the loop between helix 1 and 2 in Im9 is only 5 residues long, rather than 6 as in Im7. The notion that stability of the intermediate partly depends on loop flexibility was experimentally confirmed in Im7 (*Capaldi et al., 2002*) as well as in Im9 (*C. Friel and S. Radford, unpublished data*). Additionally, examination of a sequence alignment of Im7 and Im9 reveals that the residues that are involved in the non-native hydrophobic interactions in the intermediate of Im7 are less hydrophobic in Im9. Substituting these residues with residues that have a larger more hydrophobic side-chain does not only stabilize the

intermediate of Im9 by 10.2 kJ mol^{-1} relative to the intermediate observed in Im9 wild type but also introduces significant hyperfluorescence (*C. Friel and S. Radford, unpublished data*).

2.9. Objectives of the Presented Protein Folding Study

2.9.1. A Unified Model for Protein Folding

There is a long standing debate about the mechanism of protein folding. The diffusion-collision model states that collisions between nascent secondary structure elements direct the formation of tertiary contacts and lead to the native conformation (*Karplus and Weaver, 1976; Karplus and Weaver, 1994*). Thus the fluctuating secondary structure observed in the unfolded state of many proteins is of special importance for their folding properties. The nucleation-condensation model, however, proposes that secondary and tertiary structure is formed in parallel as the folding reaction progresses (*Fersht et al., 1992; Itzhaki et al., 1995; Matouschek et al., 1992; Nolting et al., 1997*). In this case the fluctuating secondary structure and hydrophobic clusters that characterize the unfolded state are of similar importance for folding. Recently a unified model of folding was presented, which postulates that the folding mechanism is modulated by the intrinsic conformational preference for secondary structure and the hydrophobicity of a sequence (*Daggett and Fersht, 2003; Gianni et al., 2003*). The authors present evidence that the folding mechanism of a group of homologous proteins becomes more diffusion-collision-like when the secondary structure propensity of the sequence increases.

2.9.2. Applying the Unified Model for Protein Folding to Im9

The objective of the presented study is to elucidate the effect of the intrinsic stability of helices and of hydrophobicity on the folding mechanism of a natural protein. Site-directed mutagenesis was used to design two groups of mutants. In the first group the intrinsic stability of each of the three long helices of Im9 was increased. In the second group hydrophobic residues were introduced on the solvent exposed surface of each of the three long helices. Thus, the role played in folding by hydrophobic residues which do not participate in the hydrophobic core was investigated.

We decided to use Im9 as model system for this study because it is a small, helical and very well characterized protein (*Ferguson et al., 1999; Friel et al., 2003; Gorski et al., 2001*). Its folding mechanism depends on the buffer conditions: at neutral pH it folds in a two-state

manner, at acidic pH and in the presence of Na_2SO_4 to the reaction buffer a kinetic intermediate is observed *(Gorski et al., 2001)*. Therefore, meaningful comparisons and conclusions are possible regardless of whether the designed proteins fold in a two-state or a three-state manner.

Interestingly, a second member of the group of immunity proteins, Im7, always folds through a kinetic intermediate. This is especially intriguing in the context of this study: Firstly, in Im7 the sequence of helix I shows a markedly higher propensity for secondary structure than Im9; Secondly, the sequence comprising helix II in Im7 is more hydrophobic than the equivalent sequence in Im9. The presented study thus also investigates whether the different folding properties of the highly homologous and topologically nearly identical proteins Im7 and Im9 are due to their differences in helical propensity and hydrophobicity.

2.9.3. Main Findings of the Presented Study on Im9

Increasing the helical propensity of sequence segments that adopt helical structure in Im9 often increases the stability of the intermediate, transition and native state. The extent of the stabilization depends upon which of the three long helices is stabilized. A small stability increase correlates with a significant compaction of the unfolded state. We speculate that the high helical propensity of a sequence leads to helical structure in the unfolded state. When this occurs, the stabilization of the intermediate, transition and native state is less pronounced because the unfolded state becomes not only structured, but stabilized as well, which prevents acceleration of the folding reaction. However the helices showing the highest helix propensity after the mutations are not those that become formed in the unfolded state, implying that other factors also define the properties of the unfolded state. When no significant compaction of the unfolded state is observed the folding rate in water is increased, indicating that folding is facilitated by increasing helical propensity.

Increasing the hydrophobicity of the solvent exposed faces of the helix I and IV in Im9 leads to a destabilization of the native and transition states according to the reverse hydrophobic effect. In contrast, increasing the hydrophobicity of the solvent exposed face of helix II stabilizes the intermediate, transition and native states. The intermediate state shows a very slight hyperfluorescence. In these mutants there are no apparent effects on the unfolded state. It is thus difficult to generalize the role of hydrophobicity in Im9 folding, especially when the concerned residues are fully solvent exposed. In fact, the role of hydrophobicity in folding

depends on the exact role of the considered residue in the folding mechanism of the respective protein. Generally speaking, our results show that the hydrophobicity of surface exposed residues can play an important role in determining the folding mechanism. Specifically, we show that the surface of helix II plays an important role in determining the folding properties of Im9.

Finally, we would like to point out that the differences in the folding mechanism of Im9 and Im7 can not be completely rationalized by differences in surface hydrophobicity and helical propensity. Some of the Im9 derivatives with increased helical propensity or increased hydrophobicity fold through a kinetic intermediate. This intermediate, however, is never as stable as the intermediate observed in Im7 nor does it show the hyperfluorescence that is typical of the Im7 intermediate. Moreover, the stabilization of the intermediate in Im9 is usually accompanied by a similar stabilization of the native state whereas the native state of Im7 is much less stable than that of Im9. We also note that increasing the helical propensity of helix I, the helix that shows the highest helical propensity in Im7, has virtually no effect on the stability of the intermediate of Im9.

3. MANUSCRIPT: "HELIX STABILITY AND HYDROPHOBICITY IN THE FOLDING MECHANISM OF THE BACTERIAL IMMUNITY PROTEIN IM9."

Helix Stability and Hydrophobicity in the Folding Mechanism of the Bacterial Immunity Protein Im9

Susanne Cranz[1*], Claire T. Friel[1] and Sheena E. Radford[1]

[1]School of Biochemistry and Microbiology, University of Leeds, Leeds LS2 9JT, UK.

[*]Current address: University of Zurich, Biochemical Institute, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.

Correspondence should be addressed to S.E.R.

Telephone: 0113 343 3170; Fax: 0113 343 3167; email: s.e.radford@leeds.ac.uk

Abstract

Recently a unifying model of protein folding was proposed where the ratio of secondary structure stability and hydrophobicity modulates the folding mechanism. Here we test the influence of hydrophobicity and inherent stability of secondary structure on the folding properties of the well-characterized all-helical protein Im9* by individually stabilizing the three long helices formed in the transition state of Im9*. In all cases a previously hidden intermediate becomes observable when helices are stabilized by mutation of solvent exposed residues. Interestingly, the homologous Im7 protein folds with an observable intermediate and features a high helical propensity in the sequence comprising helix I. Detailed analysis reveals different effects upon stabilizing different helices in Im9: this is related to changes in structure and energy of the unfolded state. Increasing the hydrophobicity of individual helices by mutating solvent exposed residues generally causes a reverse hydrophobic effect and the previously observable intermediate becomes hidden. However, increasing hydrophobicity of helix II in Im9 stabilizes the intermediate, transition and native states. We note that in the aforementioned Im7 protein helix II is more hydrophobic as well, confirming the special role of helix II in the folding pathway of immunity proteins.

Keywords: Folding, immunity proteins, helical propensity, intermediate, chevron plot

Introduction

The role of α-helical propensity and hydrophobicity in determining the rates and mechanisms of protein folding has been explored by a number of theoretical and experimental studies (*Bai, 1999; Chowdhury et al., 2003; Daggett and Fersht, 2003; Gong et al., 2003; Lopez-Hernandez et al., 1997; Munoz et al., 1996; Munoz et al., 1994; Poso et al., 2000; Uversky, 2002; Viguera et al., 2002; Viguera et al., 1997; Villegas et al., 1996; Villegas et al., 1995*). These studies have shown that the rate of folding of proteins that fold with simple two-state-kinetics is dependent upon their content of secondary structure and its stability. This supports the diffusion-collision model of protein folding, which describes a hierarchical process in which pre-formed elements of secondary structure with minimal stability form intermediates of increasing complexity and stability resulting ultimately in the native conformation (*Baldwin and Rose, 1999i; Baldwin and Rose, 1999ii*). It also predicts the formation of stable non-native intermediates either involving non-native pairing of folding elements or the influence of non-native secondary structure propensity of the sequence on the folding mechanism (*Beck et al., 2001*). In other models hydrophobic collapse dominates the early stages of folding, causing compaction of the protein and thereby defining geometric constraints in which the conformational search for the native state can take place (*Rose and Roy, 1980*). The nucleation-condensation model proposes that secondary and tertiary structure are stabilized concomitantly, with most, if not all, residues contributing towards the stability of the folding nucleus which characterizes the rate-limiting transition state (TS) (*Otzen et al., 1994*). Which model dominates folding for a particular sequence is in fact determined by a fine balance between the intrinsic stability of secondary structural elements and the propensity of the polypeptide chain to undergo hydrophobic collapse (*Bennion and Daggett, 2003; Gianni et al., 2003*). Thus folding actually involves a continuum of models where the balance between the content of secondary structure and its stability and hydrophobicity determines which mechanism dominates the folding reaction.

A number of recent studies on the role of helix stability in protein folding have shown that increasing the intrinsic stability of helices by substitution of solvent exposed residues results in the global stabilization of the native protein. For example, increasing the helical propensity of natural α-helices in the activation domain of human procarboxypeptidase A2 (ADA2h) results in a stability increase based on an acceleration of the refolding rate constant or a decrease in the rate of unfolding (*Viguera et al., 1997*). The former case has been interpreted to mean that the

affected helix is formed in the transition state but not in the unfolded state, whilst the latter indicates that the helix is fully formed only after the rate-limiting transition state has been traversed. A similar study conducted on the protein CheY resulted in a much smaller increase in protein stability than predicted based on the stabilization of individual helices, presumably because the helices are also formed in the denatured state (*Munoz et al., 1996*). When effects on all states along the reaction coordinate are taken into consideration engineering helical propensity is a powerful method of elucidating the role of individual secondary structural elements in the mechanism of folding,

Whilst the importance of hydrophobic collapse in folding has been appreciated for decades (*Dill, 1990*) predicting the effect of altering the hydrophobicity on the mechanism of protein folding is also complex, since the role of a certain hydrophobic residue is determined by its individual role in the folding reaction. For example, decreasing the size or hydrophobicity of residues involved in the formation of the folding nucleus and in the core of the native state can slow folding and decrease protein stability (*Northey et al., 2002; Richards and Lim, 1993*). By contrast, increasing the size of hydrophobic side-chains within the core that are not involved in the folding nucleus can selectively stabilize the transition state relative to the native state (*Northey et al., 2002*). Increasing the hydrophobicity of residues that are solvent exposed in the native state can also stabilize the unfolded state by facilitating the formation of non-native hydrophobic clusters, resulting in the so-called inverse hydrophobic effect (*Munoz et al., 1994; Pakula and Sauer, 1990*). The role of hydrophobicity in the mechanism of folding, therefore, depends on the precise role of the considered residue in the mechanism of folding of that particular protein.

Here we investigate the role of helical propensity and hydrophobicity in tailoring the kinetic folding mechanism of the immunity protein, Im9. Previous results have shown that this four helix protein (Figure 1) folds with a two-state mechanism at pH 7.0 and 10 ºC, whilst at lower pH and in the presence of 0.4 M Na_2SO_4 an intermediate becomes populated during folding (*Ferguson et al., 1999*). Interestingly, the Im9 homologue, Im7, that is 60% identical in sequence to Im9 (*Friel et al., 2003*) and has a similar four helical structure (*Ferguson et al., 1999*), also folds with a three-state mechanism involving the population of a stable intermediate (6.8 kJ mol^{-1} at 10 ºC, pH 7.0) that contains three of the four native helices (helices I, II and IV) packed around a specific hydrophobic core (*Capaldi et al., 2002*). The intermediate has been shown to be on-pathway (*Capaldi et al., 2001*) and is stabilized by both native and non-native

interactions (*Capaldi et al., 2002*). The rate-determining step in the folding then involves the disruption of non-native interactions so as to allow helix III to dock onto the developing structure stabilizing the protein in the native conformation (*Capaldi et al., 2002*). Using rational redesign of the Im9 sequence we have recently demonstrated that Im9 forms a similar intermediate state during folding and that this species is simply too unstable to detect using ensemble methods (*C. Friel and S. Radford, unpublished data*). These structural homologues thus fold with similar structural mechanisms despite the apparent differences in their kinetic folding mechanisms at pH7.0. One consequence of the sequence divergence in Im7 and Im9 is that these proteins have different helix propensity as judged by the helix/coil transition algorithm AGADIR (*Lacroix et al., 1998*). These proteins differ most significantly in sequence in the N-terminal half of the polypeptide chain, specifically in the regions encompassing the C-terminal residues of helix I, the loop connecting helices I and II and in helix II (Figure 1c), since residues in these regions play an important functional role in determining the specificity of each immunity protein for its cognate colicin toxin (*Dennis et al., 1998*). The helix propensity of helix I is significantly greater for Im7 than for Im9 (Figure 2A). The two sequences also show subtle differences in hydrophobicity (Figure 2D); the C-terminal region of helix I is more hydrophobic in Im9, whilst especially the N-terminal part of helix II is more hydrophobic in Im7. This pair of proteins provides an ideal system for dissecting the role of sequence, secondary structure propensity and hydrophobicity in determining the kinetic mechanism of folding.

Materials and Methods

Mutagenesis, protein expression and purification

Site directed mutations of Im9* (a hexa-histidine tagged version of the wild-type protein (*Gorski et al., 2001*) were made using the QuikChange site-directed mutagenesis kit (Stratagene) using a template carrying the Im9* gene in pTrc99a (*Gorski et al., 2001*). The proteins were overexpressed and purified as described previously (*Gorski et al., 2001*). The residue numbers used in this paper are those of the untagged protein. Every mutant was sequenced to ascertain that the gene had only the desired substitutions. All proteins studied were confirmed to be within 1Da of the expected mass as analysed by electrospray ionization mass spectrometry (ESI-MS) and ≥ 95% pure as determined by SDS-PAGE. The extinction coefficient of each protein was determined using the method of Gill and Von Hippel (*Gill and von Hippel, 1989*) at pH 6.0 and were found to be 9442, 11376, 8560, 9528, 9721, 8930 $M^{-1}cm^{-1}$, for Im9 WT H1, H2, H4, H12, H2P, respectively.

Data collection

All folding and unfolding experiments were performed as described previously using an Applied Photophysics SX18.MV stopped flow fluorimeter (*Ferguson et al., 1999*). Briefly, for refolding experiments the protein (47 μM) denatured in 8 M urea was refolded by 1:10 (v/v) ratio into buffer A (50 mM sodium phosphate, pH 7.0 or 6.0 containing 2 mM DTT and 1 mM EDTA). All experiments were performed at 10 ºC. Refolding time-courses were well described by a double exponential function, where the slower second phase comprising 10% of the amplitude was attributed to folding events limited by proline isomerization (*Ferguson et al., 1999*). In the mutant H4 this slower rate of folding was not observed. When the stability of an intermediate was very low, additional refolding experiments were performed in which 47 μM protein in 5-7 M urea was diluted in a 1:10 (v/v) ratio into buffer A resulting in final urea concentrations of 0.45 M to 0.63 M. Folding rate constants for the protein H12 at pH 7.0 were measured from 1.25 M to 8.0 M final urea concentration, since at lower urea concentrations the protein refolds too rapidly to be accurately measured using stopped flow. For unfolding experiments native protein (47 μM) in buffer A was diluted 1:10 (v/v) into buffers containing different concentrations of urea. In all cases unfolding followed a single exponential function.

Data analysis

The observed rate constants and signals for each data set were fitted simultaneously to either a two-state model (Scheme 1) or a three-state model involving an on-pathway intermediate (Scheme 2) as discussed previously (*Capaldi et al., 2002; Capaldi et al., 2001; Ferguson et al., 1999*).

$$U \underset{k_{nu}}{\overset{k_{un}}{\rightleftharpoons}} N$$

Scheme 1

$$U \underset{k_{iu}}{\overset{k_{ui}}{\rightleftharpoons}} I \underset{k_{ni}}{\overset{k_{in}}{\rightleftharpoons}} N$$

Scheme 2

To constrain the fits for the mutants H1, H2, H12 and H4 at pH 7.0 in which the intermediate formed during folding is rather unstable, the unfolding slope m_{UI} was fixed to the value obtained for WT-type Im9* (-0.42 kJ $mol^{-1}M^{-1}$). At pH 6.0 the unfolding slope m_{UI} was fixed to -0.38 kJ $M^{-1}mol^{-1}$, which is the arithmetic mean of the m_{UI} values measured at pH6.0 with Im9 WT, H1, H2, H4 and H2P. The standard deviation of the averaged values is 0.06 kJ $M^{-1}mol^{-1}$. The energy of the TS as it appears in Figure 6 A and B was calculated using equation (1) where 4.8×10^8 is the pre-exponential factor:

$$\Delta G_{TS} = -RT \ln \frac{k_{IN}}{4.8\times10^8} \qquad (1)$$

The energies of TS2 (Transition state 2) as they appear in Figure 4 and 6 were determined using equation 2 where $4.8\times10^8$is the pre-exponential factor:

$$\Delta G_{TS2} = -RT \ln K_{UI} - RT \ln \frac{k_{IN}}{4.8\times10^8} \qquad (2)$$

Φ_I and Φ_{TS}-values were calculated according to equations 3 and 4:

$$\Phi_{TS} = \frac{\Delta\Delta G_{UTS}^{wt-mut}}{\Delta\Delta G_{UN}^{wt-mut}} = \frac{\Delta G_{UTS}^{wt} - \Delta G_{UTS}^{mut}}{\Delta G_{UN}^{wt} - \Delta G_{UN}^{mut}} \qquad (3)$$

$$\Phi_I = \frac{\Delta\Delta G_{UI}^{wt-mut}}{\Delta\Delta G_{UN}^{wt-mut}} = \frac{\Delta G_{UI}^{wt} - \Delta G_{UI}^{mut}}{\Delta G_{UN}^{wt} - \Delta G_{UN}^{mut}} \qquad (4)$$

where $\Delta\Delta G_{UI}$ is the change in free energy of the intermediate, $\Delta\Delta G_{UTS}$ is the change in free energy of the rate-limiting transition state and $\Delta\Delta G_{UN}$ is the change in free energy of the native state, and wt and mut correspond to the wild-type and mutant proteins, respectively. Errors on all constants were estimated by varying the parameters as widely as possible and determining

the range of values that still adequately describe the data. Errors on ΔG, equilibrium m-values and β-values were calculated using the common mathematical procedures for error propagation. When errors were asymmetric, the larger error is shown. The errors on the Φ-values were determined by calculating the smallest and the largest Φ-values possible considering the errors on the energies of the native, intermediate and transition states. For Φ-values, the asymmetric errors are shown as such.

The fluorescence of the intermediate and native state was normalized to the fluorescence of the unfolded state in 8M urea.

Results

Designing Helix Propensity

To investigate the role of helix propensity in Im9* folding, three mutants of Im9* were created in each of which one of the three long helices, helices I, II or IV, was stabilized by substitution of one or more solvent exposed residues, designed using the program AGADIR (*Lacroix et al., 1998*). Helix III is not structured in the early folding intermediate in Im7 and is not formed in Im9 until after the rate limiting transition state is traversed (*Capaldi et al., 2002; Friel et al., 2003*). This sequence, therefore, was not mutated in this study. The mutants containing stabilizing substitutions in helices I, II and IV are referred to as H1, H2 and H4, respectively. A fourth mutant, named H12, was also created in which both helices I and II are stabilized. In all mutants changes in the overall hydrophobicity of the sequence were minimized (Figure 2B, E). Mutant H1 was obtained by changing the solvent exposed residues, Q17 and T20 in Helix I, to Ala (Figure 1A). Additionally, two mutations were introduced that stabilize the helix through interactions with the dipole (A13E and T21R). Although Glu has a lower intrinsic preference for helical dihedral angles than Ala, it forms interactions with the helix dipole when situated toward the amino-terminus of a helix and also reduces the increase in hydrophobicity caused by substitution of residues 17 and 20 with Ala. The mutation T21R also increases helix propensity because the intrinsic preference of Arg for helical dihedral angles is greater than that of Thr. This residue also stabilizes the helix through interactions with the helix dipole. In the resulting quadruple mutant, A13E, Q17A, T20A and T21R, the helix-propensity of helix I is increased by ~15% whilst its overall hydrophobicity remains unaltered (Figures 2B, E). In the mutant H2, two solvent-exposed residues, V34 and T38, were changed to Ala (Figure 1A, C). These substitutions increase the helical propensity of helix II by approximately 25% without changing its hydrophobicity (Figures 2B, E). Increasing the helical propensity of helix IV was more difficult, since several surface exposed residues in helix IV, such as K72, Q73, A76 and A77 (Figure 1C) are already occupied by residues that show a high helical propensity. In addition, we preferred not to alter the N-terminal capping position since it is occupied by proline that may play an important role in defining the limits of the helix. The carboxy-terminal capping position is occupied by glycine, mutation of which is known to cause ground state effects. In addition, the single tryptophan residue (W74) could not be altered since it provides the fluorescent probe of folding. To create the mutant, H4, therefore, four mutations were made (Figure 1 A,C). N69 and T70 were both replaced with Ala, whilst the mutations S65D and Q73R were included to stabilize the helix by interactions with the helix dipole. Of these

substitutions, mutation T70A is almost entirely responsible for the observed increase in helical propensity of ~10% (Figure 1C). However threonine 70 is not entirely solvent exposed in native Im9, its β-CH_3 group forming a van-der-Waals contact to Leu52 in helix III. Changing threonine to alanine abolishes this interaction and thus is expected to destabilize the native state. Because helix III is not structured in the intermediate we predict that the stability of the intermediate should not be affected by this substitution.

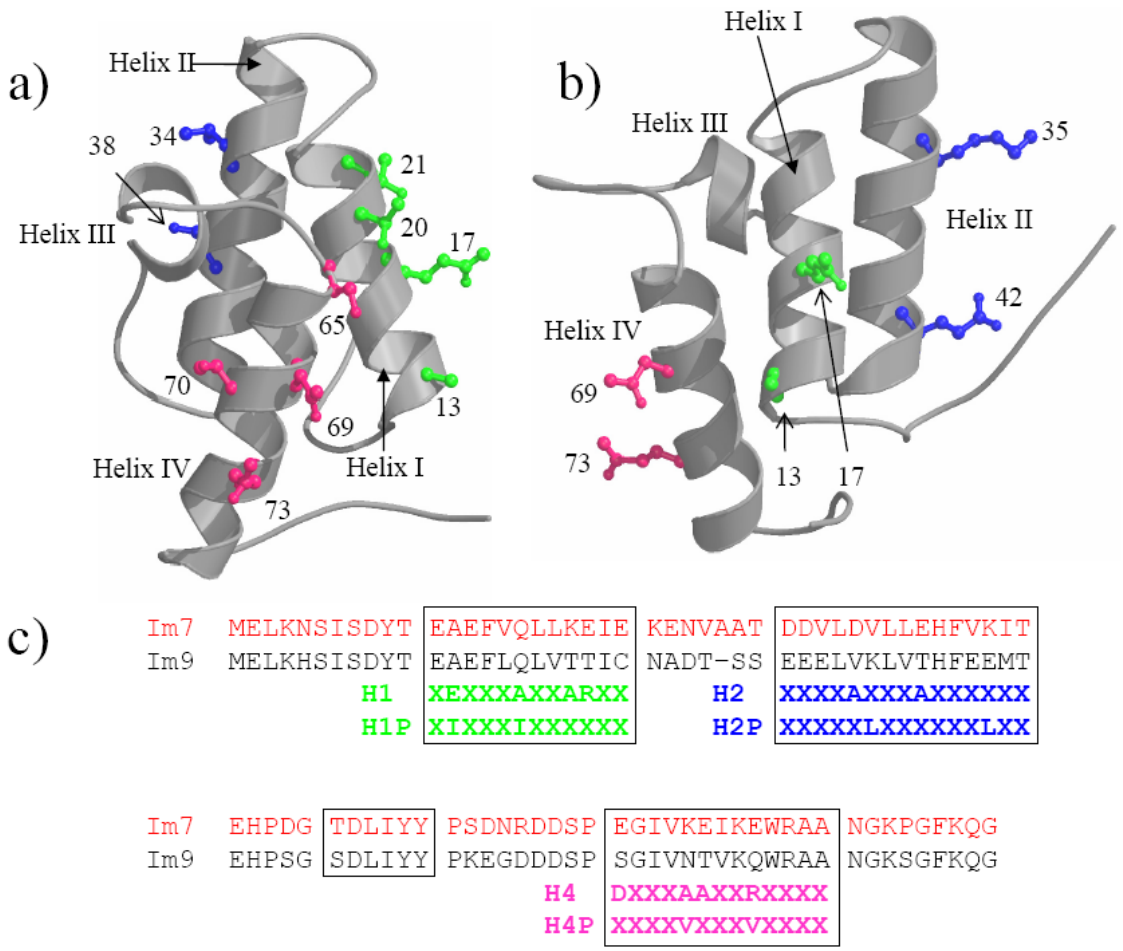


Figure 1: Ribbon diagrams of Im9 illustrating its four-helix structure and the positions of residues mutated in this study. The images were prepared using MOLSCRIPT (*Kraulis, 1991*) and Raster3D (*Merrit, 1997*) using the co-ordinates from the solution structure of Im9 (1IMQ) (*Osborne et al., 1996*). **A)** Side chains mutated to engineer helix propensity. Residues shown in green were changed to obtain the mutant H1 (A13E, Q17A, T20A T21R); side chains shown in blue were changed in the mutant H2 (V34A, T38A); and side chains shown in pink were changed to create the mutant H4 (S65D, N69A, T70A, Q73R). **B)** Side chains mutated to alter the hydrophobicity of different helices. The residues shown in green were changed in mutant H1P (A13I, Q17I); side chains shown in blue were changed in the mutant H2P (K35L, E42L) and residues shown in pink were changed in the mutant H4P (N69V, Q73V). **C)** Alignment of the sequences of Im7 and Im9. The position of the four helices is shown in boxes. Mutations introduced to engineer helix propensity and hydrophobicity of Im9 are indicated.

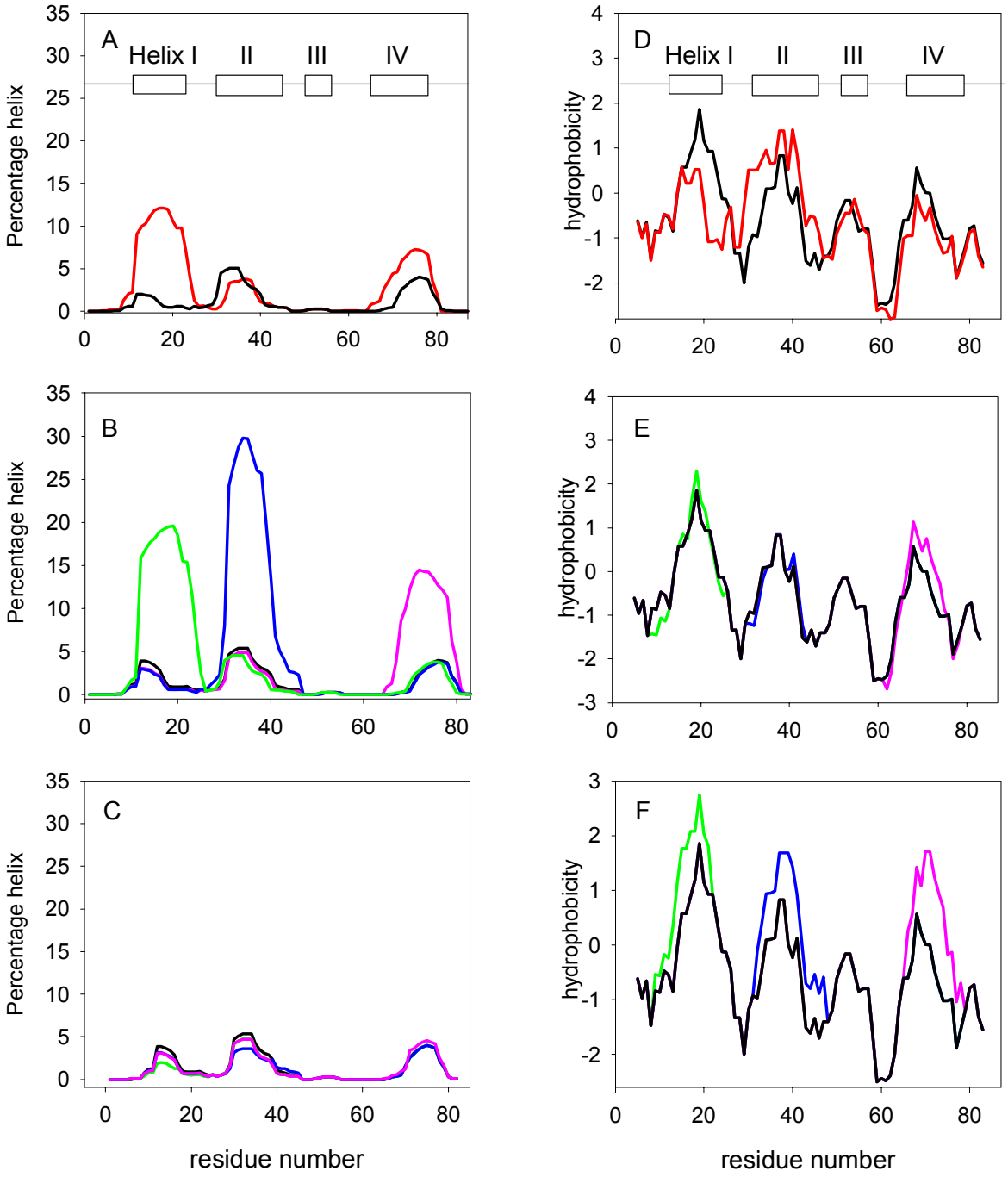

A
Helix I
II
III
IV
Percentage helix
B
C
residue number
D
hydrophobicity
E
F
residue number

Figure 2: A) AGADIR (*Lacroix et al., 1998*) prediction of the fractional helix population of the sequence of Im7 (shown in red) and Im9 (black) as a function of residue number. **B)** AGADIR prediction of the fractional helix population of Im9 wild-type (WT) (black) and the mutants H1 (green), H2 (blue) and H4 (pink). **C)** AGADIR prediction of the fractional helix population of wild-type Im9 (black) and the mutants H1P (green), H2P (blue), H4P (pink). All predictions were determined at pH 7.0, ionic strength 1 and 283 K. **D)** Hydrophobicity plot of Im7 (red) and Im9* (black). **E)** Hydrophobicity plot of wild-type Im9 (black) and the mutants H1 (green), H2 (blue), and H4 (pink). **F)** Hydrophobicity plot of wild-type Im9 (black) and the mutants H1P (green), H2P (blue) and HP4 (pink). The hydrophobicity plots were obtained by the program ProtScale provided by Expasy Molecular Biology server (www.expasy.org). The hydrophobicity scale of Kite and Doolittle (http://ca.expasy.org/tools/pscale/Hphob.Doolittle.html) was used with a window size of 7. Note that in Figure 1A and 1D a gap was left between residues 27 and 28 since Im9 is one residue shorter than Im7 (Figure 1C).

Effect of Increased Helix Propensity on the Folding Kinetics of Im9

All of the mutants were cloned and over-expressed to high levels (see Methods). To ensure that all of the variants had folded correctly to a native-like structure, each protein was examined using far-UV CD and fluorescence. All of the proteins had a helical content similar to that of wild-type Im9* as judged by far-UV CD, only H1 showed a small (<10%) reduction in helicity (data not shown). In addition, the fluorescence emission of the single tryptophan in all of the proteins is highly quenched (data not shown), presumably by the docking of Trp74 against His46 in the core of the proteins, as observed for the wild-type protein (*Ferguson et al., 1999*). Together these data demonstrate that the secondary and tertiary structure of the variants is not changed substantially relative to wild-type Im9*.

The folding and unfolding kinetics of wild-type Im9* at pH 7.0 and 10 °C are shown in Figure 3A. The data show that the logarithm of the folding and unfolding rate constants of the protein depend linearly on the denaturant concentration over the range studied, demonstrating, in accord with previous results (*Ferguson et al., 1999; Gorski et al., 2001*) (*Friel et al., 2003*), that Im9* folds with two-state kinetics under these conditions. In addition, there is no evidence for a burst-phase change in signal in the dead time (3 ms) of refolding (data not shown) and the ΔG_{UN} and total M-value (M_{UN}) obtained by kinetic experiments (27.1±0.8 kJ mol^{-1} and 5.3±0.1 kJ $mol^{-1}M^{-1}$, respectively, Table 1) agree closely with the same parameters obtained using equilibrium conditions (26.1±0.5 kJ mol^{-1} and 4.62±0.09 kJ $mol^{-1}M^{-1}$, respectively (*Ferguson*

et al., 1999)), confirming that a two-state mechanism describes the folding and unfolding of Im9* under these conditions. By contrast, at pH 6.0, the folding kinetics of wild-type Im9* show a non-linear dependence of the logarithm of the folding rate constant *versus* urea-concentration (Figure 3B). These data also accord with previous results obtained at low pH in the presence of 0.4 M Na_2SO_4 (*Gorski et al., 2001*), demonstrating that under mildly acidic conditions Im9* populates an intermediate in the dead-time of folding. Previous studies involving global analysis of the pH-dependence of the folding kinetics of Im9 (*Gorski et al., 2001*), together with detailed analysis of the folding kinetics of a series of Im9* variants that fold with three-state kinetics involving the formation of a hyper-fluorescent intermediate in the burst phase, (*C. Friel and S. Radford, unpublished data*), have ruled out alternative models for the folding of Im9*, involving movement of the rate-determining transition state or transient aggregation. The refolding kinetics of all of the Im9* variants that fold with three-state kinetics created here were thus fitted to an on-pathway, three-state model (see Methods).

As shown in Figure 3A, increasing the helical propensity of helices I, II and IV has a dramatic effect on the folding kinetics of Im9* at pH 7.0. By contrast with the wild-type protein, all of the variants fold with clear three-state kinetics at pH 7.0 and in no case was a change in signal in the burst phase observed demonstrating that, akin to the intermediate formed during folding of wild-type Im9* (*Gorski et al., 2001*) the intermediate formed during the folding of H1, H2, H12 and H4 has a fluorescence signal similar to that of the denatured protein (data not shown). In accord with the view that the intermediate formed during Im9* folding is a three helical species (*C. Friel and S. Radford, unpublished data*), these results demonstrate that increasing the helical propensity of any of the three long helices in Im9* stabilizes the intermediate formed during folding such that it becomes significantly populated (Figure 3A). Importantly, control experiments in which the refolding kinetics of H4 and H12 were measured as a function of protein concentration (88-0.3 μM for H4 in 1.0 M urea at pH 6.0 and 85-0.33 μM for H12 at 1.25 M final urea concentration at pH 6.0) showed no dependence of the rate constant of folding on protein concentration, demonstrating that the increased stability of the Im9* folding intermediate cannot be attributed to transient intermolecular association.

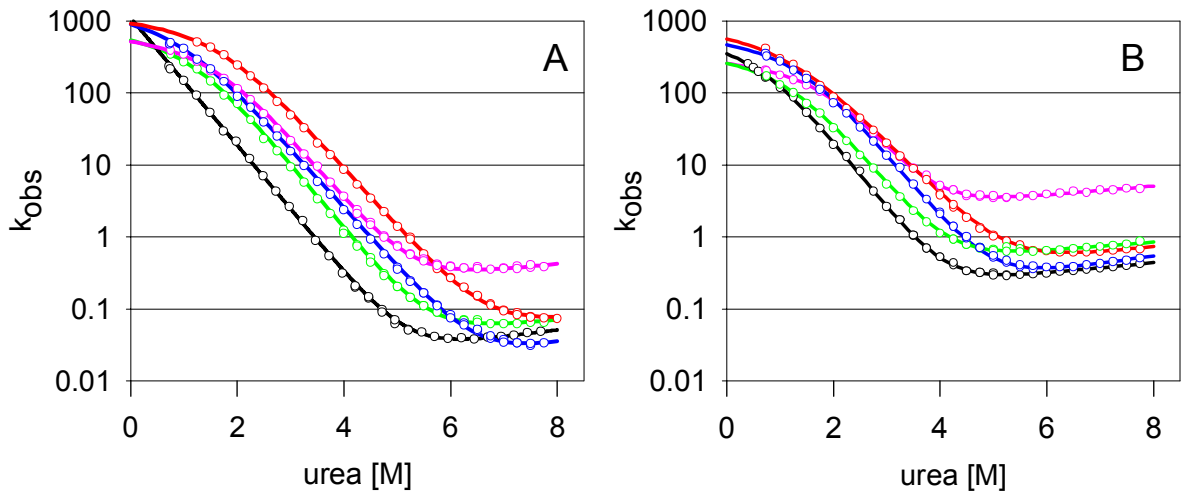


Figure 3: Denaturant-dependence of the folding and unfolding kinetics of the Im9* mutants H1 (green), H2 (blue), H4 (pink) and H12 (red) at pH 7.0 (A) and pH 6.0 (B) at 10 °C. The continuous lines show the best fit of the data to either a two-state model (wild-type Im9* at pH 7.0) or to a three-state model involving an on-pathway intermediate (all other data).

Consistent with the data obtained at pH 7.0, all four variants with increased helix propensity also show three-state folding kinetics at pH 6.0 (Figure 3B). By contrast with the results at pH 7.0, however, the effect of the mutations can be readily quantified at the lower pH, since wild-type Im9* also folds with three-state kinetics at this pH (Table 1). Fitting the folding and unfolding kinetics for all proteins at pH 6.0 to a three-state model thus allows the effect of the mutations on the stability of the intermediate and native states to be quantified, whilst changes in the m-values can be used to determine whether the new sequences affect the compactness of different states along the reaction coordinate. These results are summarized in Figure 4 and Table 1 and are described in more detail below.

The data shown in Figure 4 demonstrates that stabilizing different helices in Im9* changes the folding kinetics in subtly different ways. For example, the stability of the intermediate formed during the folding of H1 at pH 6.0 is virtually unchanged relative to its counterpart in wild-type Im9* ($\Delta\Delta G_{UI}$ = -0.4 kJ mol^{-1}), whilst for the variants H2 and H4 a significant increase in the stability of the intermediate is observed ($\Delta\Delta G_{UI}$= -2.8 kJ mol^{-1} and -4.9 kJ mol^{-1}). Interestingly, the sequence encompassing the native helix I differs most significantly in helix propensity in Im7 and Im9; Im7 having a much higher helical propensity in this region (Figure 2A). The data suggest that the difference in helix propensity in this region alone cannot account

for the different folding kinetics of Im7 and Im9 at neutral pH. In addition, changing the helical propensity of helix I has no effect on the stability of TS2 and the stability of the native state. Thus, the stabilization of 8.9 kJ mol^{-1} predicted (using AGADIR (*Munoz et al., 1994*)) to occur upon stabilization of this helix by the mutations introduced is not observed. The simplest explanation for this is that the mutations stabilize helix I in the unfolded state to the same extent as in the intermediate and TS2 such that no net stabilization for the structured species is observed. In support of this suggestion both M_I and M_{UN} are significantly reduced in H1 relative to wild-type Im9*, suggesting that the unfolded state in H1 is more compact than its wild-type counterpart (Table 1).

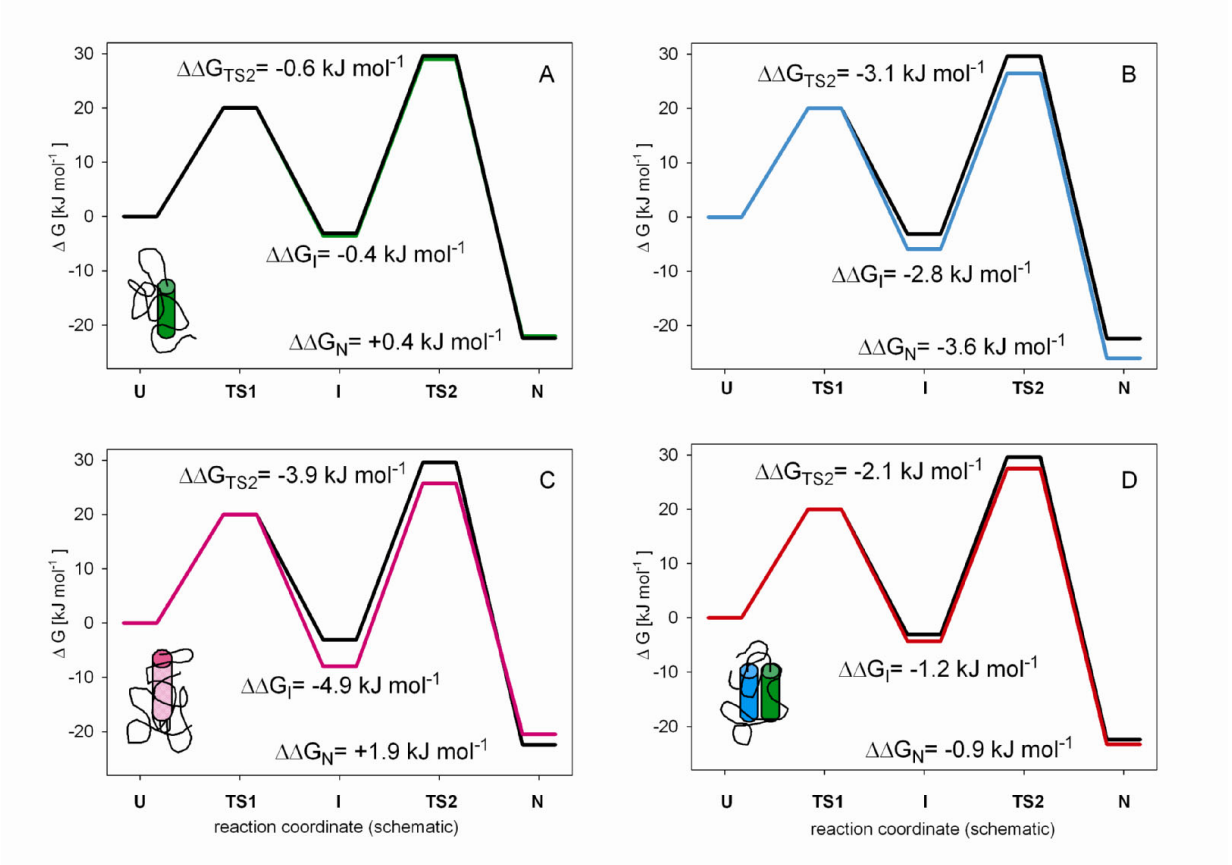


Figure 4: Free energy diagrams illustrating the effect of stabilizing different helices on the folding of Im9*: A) H1 (green), B) H2 (blue), C) H4 (pink) and D) H12 (red) at pH 6.0. The data for the wild-type protein are shown in black on each panel. When the data for a mutant can be explained by a ground-state effect, a schematic depiction of the changed ground-state is shown on the respective energy diagram. The free energies are plotted in kJ mol^{-1}. The energy of TS1 (Transition state 1) was not determined experimentally and was fixed to 20 kJ mol^{-1}. The energy of TS2 was determined as described in methods.

The data for H2 differ significantly from those for H1. Thus, H2 folds more rapidly than wild-type Im9* and H1 under both pH values studied, demonstrating that TS2 is stabilized relative to both the unfolded state and the populated intermediate in this variant (Figure 4B). The native state of H2 unfolds more slowly than in wild-type Im9*, suggesting that it is stabilized more than TS2. The progressive stabilizations of the intermediate, TS2 and the native state relative to the unfolded state in this variant ($\Delta\Delta G_{UI}$= -2.8 kJ mol^{-1}, $\Delta\Delta G_{UTS2}$ = -3.1 kJ mol^{-1}, $\Delta\Delta G_{UN}$ = -3.6 kJ mol^{-1} (Figure 4B), combined with the observation that the stabilization observed for each of the native state is close to that predicted using AGADIR (4.1 kJ mol^{-1}) (*Munoz and Serrano,*

1994; Munoz and Serrano, 1995) suggests that the mutations introduced in H2 do not affect the stability (or structure) of the unfolded state. Consistent with this, M_{UN} for H2 is close to the value for wild-type Im9* (Table 1).

The mutant H12 shows the largest effect on folding of all of the variants studied ((Figure 3B). The folding rate constant for this variant is increased more than 8-fold when folding is two-state (e.g. in 3.5 M urea) and 2-fold when it is three-state (e.g. in 0.75 M urea) (Figure 3B and Table 1), suggesting that the rate-limiting transition state is stabilized significantly relative to the unfolded state in this variant. Interestingly, whilst the intermediate and native states in this variant are stabilized relative to the equivalent species in wild-type Im9*, the stabilization observed is less than predicted assuming that the effects of H1 and H2 on the stability of these states is additive (expected $\Delta\Delta G_{UI}$ = -3.2 kJ mol^{-1} vs. measured $\Delta\Delta G_{UI}$ = -1.2 kJ mol^{-1}, expected $\Delta\Delta G_{UN}$ = -3.2 kJ mol^{-1} vs. measured $\Delta\Delta G_{UN}$ = -0.9 kJ mol^{-1}). In addition, M_{UN} for H12 is reduced significantly relative to wild-type Im9, (4.48 and 5.48 kJ $mol^{-1}M^{-1}$, respectively (Table 1)). Together, this suggests that increasing the helical propensity of helices I and II results in increased population of helical structure of the sequences in the unfolded state (Figure 4D), the non-additive effect observed in H12 further suggest that the helices can interact in the unfolded state.

Altering the sequence of helix IV in H4 destabilizes native Im9*, as expected since the mutation deletes a key interactions of the γCH_3 of T70 with residue L52 in helix III (see above). However, the mutations introduced stabilize the intermediate formed during folding ($\Delta\Delta G_{UI}$ = 4.9 kJ mol^{-1} (Table 1)) and increase the rate of folding significantly when folding is two-state (>6-fold in 3 M urea), whilst at low urea concentrations where folding is three-state the rate of folding is virtually unchanged (Figure 3B) . This indicates that the intermediate and TS2 are similarly stabilized relative to the unfolded state in this variant (by ~ 4.5 kJ mol^{-1} (Figure 4C)). This effect is smaller than predicted using AGADIR (~10 kJ mol^{-1}) (*Munoz et al., 1994; Munoz and Serrano, 1997*). The M_{UN} value is only slightly decreased relative to wild-type Im9* (5.2 and 5.48 kJ $mol^{-1}M^{-1}$, respectively (Table 1)), suggesting that the mutations do not increase the formation of helical structure in the unfolded state significantly.

Altering Surface Hydrophobicity by Design

In a second series of experiments, the effect of increasing the hydrophobicity of solvent exposed residues on the folding kinetics of Im9* was examined. Again, three mutants were created, in each of which the hydrophobicity of helix I, helix II or helix IV was increased individually (in the variants, H1P, H2P and H4P (Figure 1C)). In each case, hydrophobicity was changed by substituting solvent exposed residues with more hydrophobic side-chains without significantly influencing the helix propensity of the resulting sequence. The mutants A13I, Q17I (H1P), K35L, E42L (H2P) and N69V, Q73V (H4P) were thus created (Figures 1B,C and 2F). These mutants were designed to determine whether non-native interactions during folding, involving residues that are natively exposed, can stabilize the folding intermediate of Im9*. They also build on recent results that demonstrated that a switch from two-to three-state kinetics can be achieved in Im9* folding by increasing the hydrophobicity of buried residues (V37 and V71) or a solvent exposed residue (E41 in helix II). These sites were chosen specifically for mutation by design, based on the predicted structure of the intermediate populated during the folding of Im7* (*Capaldi et al., 2002*). Indeed, the triple mutant (V37L, E41V, V71I, named Switch Im9* (*C. Friel and S. Radford, unpublished data*)) folds with clear three-state kinetics through a highly stabilized intermediate state (ΔG_{UI} = 10.5 kJ mol^{-1}) populated in the dead-time of folding. Finally, the mutants H1P, H2P and H4P probe the importance of the difference in hydrophobicity of the sequences of helices I and II in the kinetics of Im7* and Im9* folding (Figure 2A).

Effect of Increased Surface Hydrophobicity on the Folding Kinetics of Im9*

The folding and unfolding kinetics of the mutants H1P, H2P and H4P at pH 7.0 and pH 6.0 are shown in Figures 5A and B. The data show that whilst H1P and H4P fold with two-state kinetics at both pH values studied, the variant H2P populates an intermediate in the dead-time of folding under both conditions. Comparing the data for wild-type Im9* and H2P at pH 6.0, under which conditions both proteins fold with three-state kinetics, shows that the intermediate of H2P is stabilized significantly relative to the intermediate populated during the folding of the wild-type protein (ΔG_{UI} of 3.1 kJ mol^{-1} and 7.7 kJ mol^{-1} for wild-type Im9* and H2P, respectively (Figure 6 and Table 2). Interestingly, the intermediate of H2P is more fluorescent than the native and denatured states of this protein (relative fluorescence signals in H_2O are 0.3, 0.85 and 0.63 for the native, intermediate and denatured states in water respectively), reminiscent of the fluorescence properties of the intermediate populated during the folding of wild-type Im7* (*Capaldi et al., 2001; Ferguson et al., 1999*) as well as the Im9* Switch

variants (Im9* V37L and E41V) (for example the relative fluorescence signals in water for the native, intermediate and denatured states of V37L are, 0.82 and 0.65, respectively (*C. Friel and S. Radford, unpublished data*)).

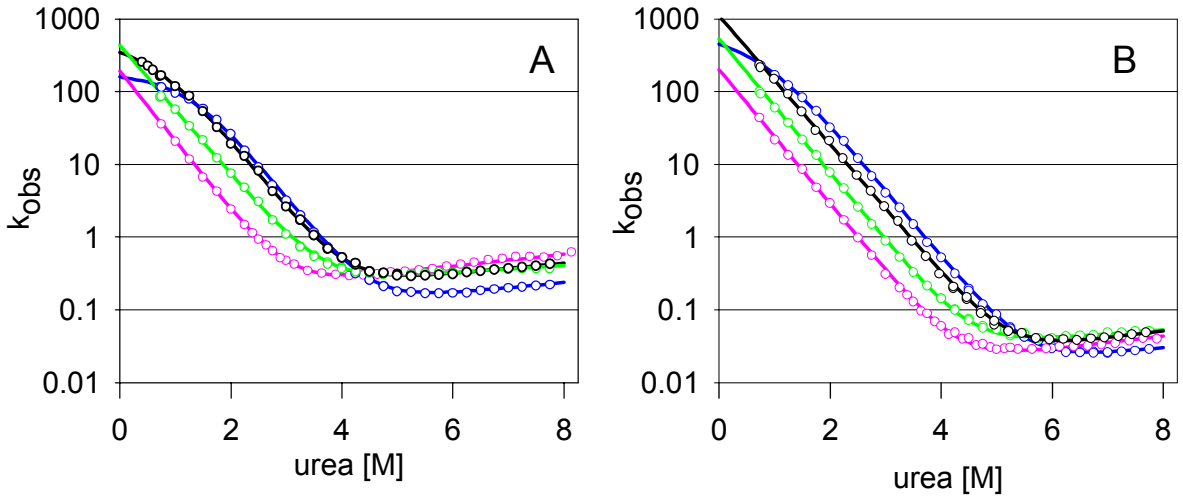


Figure 5: Denaturant-dependence of the folding and unfolding rate constants of the hydrophobicity mutants H1P (green), H2P (blue) and H4P (pink) at pH 7.0 (A) and pH 6.0 (B) at 10°C. The continuous line shows the best fit of the data to a three-state model for H2P and a two-state model for H1P and H4P. The black continuous line shows the best fit to the data for wild-type Im9*.

By contrast with the variant H2P, which is stabilized relative to wild-type Im9* ($\Delta\Delta G_{UN}$ = -3.6 kJ mol^{-1} at pH 6.0 (Table 2)), the variants H1P and H4P which fold with two-state kinetics at all pH values studied (Figure 5A, B) are destabilized significantly relative to wild-type Im9* ($\Delta\Delta G_{UN}$ is approximately 3.0 kJ mol^{-1} for both proteins at pH 7.0 (Figure 6 and Table 2)). This effect is manifested by a decreased folding rate constant, whilst the rate of unfolding is not affected by the mutations introduced, consistent with stabilization of the unfolded state presumably through a reverse hydrophobic effect (*Pakula and Sauer, 1990*) in which solvent exposed hydrophobic residues cluster in the unfolded state. Despite this, the values of M_{UN} for these variants are unchanged relative to wild-type Im9*, suggesting that the relative compactness of the unfolded and native states are not affected substantially by the amino acid substitutions introduced (Table 2).

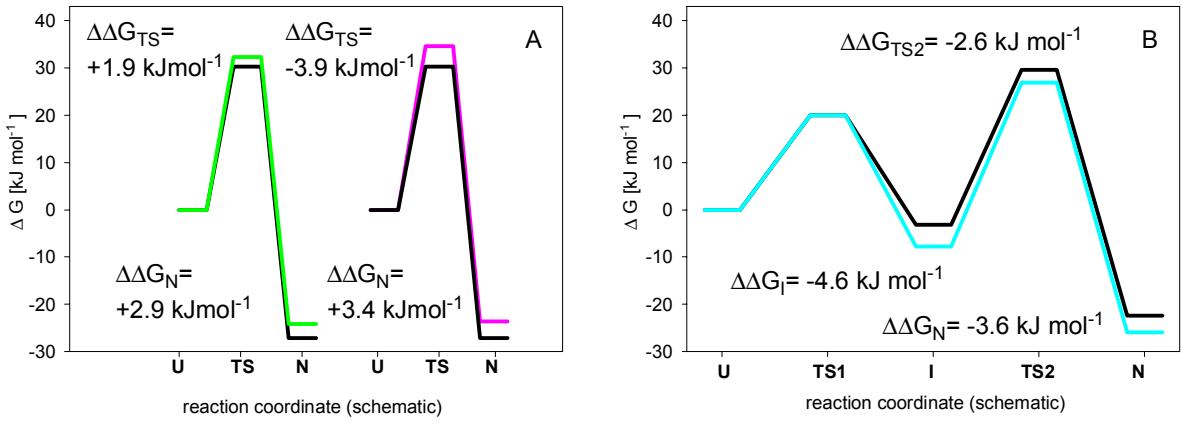


Figure 6: Free energy diagrams illustrating the effects of increasing the hydrophobicity of helices I, II and IV in the folding of Im9*: A) mutant H1P (green) mutant H4P (pink) at pH 7.0. B) Mutant H2P (blue) at pH6.0. The free energy diagram of wild-type Im9* at the respective pH is shown on each panel in black. The free energies are plotted in kJ mol^{-1}. The energy of TS1 in B) was not determined experimentally and was fixed to 20 kJ mol^{-1}. The energy of TS in A) and TS2 in B) were determined as described in methods.

In order to characterize the intermediate populated during the folding of H2P in more detail, a number of mutations were introduced and the properties of the intermediate and rate-limiting transition states determined using Φ-value analysis (*Fersht et al., 1992*). The mutations were carefully chosen based on previous analyses of the folding of Im7*, Im9* and Switch Im9* (*Capaldi et al., 2002; Friel et al., 2003; C. Friel and S. Radford, unpublished data*) to report on (i) the presence or absence of helices I, II, and IV at different stages of folding (using the mutations, L16A, V37A, I67V); (ii) whether the intermediate and rate-determining transition states lack helix III, akin to the folding of wild-type Im7*, Im9* and Switch Im9* (using the mutation I53V); and (iii) whether non-native interactions play a role in stabilizing the intermediate and rate-limiting transition states of wild-type Im9* and H2P. In parallel, ϕ-values for the same mutations were determined for wild-type Im9* under the same conditions (pH 6.0, 10 ºC) so that a direct comparison of the folding mechanisms of the two proteins as well as comparisons with wild-type Im7*, for which mutations at equivalent sites have been reported previously (*Capaldi et al., 2002*) was possible. The resulting data (Figure 7 and Table 3) show that the intermediate populated during the folding of H2P resembles that formed during the folding of wild-type Im9* at pH 6.0 in that both species contain three of the four native helices (I, II and IV), but lack helix III. Indeed, the Φ_I-values are within error identical for these

proteins over all mutations studied. Notably, Φ_I and Φ_{TS} for the mutation I53V are close to zero for both proteins, demonstrating that, akin to Im7* and Switch Im9* (*Capaldi et al., 2002*; *C. Friel and S. Radford, unpublished data*), helix III docks into the developing hydrophobic core structure only after the rate-determining step has been traversed. The absence of helix III until very late in folding is thus a common theme in the folding of all immunity proteins studied to date *(Capaldi et al., 2002; Friel et al., 2003; C. Friel and S. Radford, unpublished data*). Another common feature of immunity protein folding is the development of significant stabilizing non-native interactions that stabilize the intermediates formed early during folding (*Capaldi et al., 2002; C. Friel and S. Radford, unpublished data*). The mutations L37A and I67V probe for such events since in Im7*, these residues play a greater role in stabilizing the intermediate than in stabilizing either the rate-limiting transition state or the final native state (*Capaldi et al., 2002; C. Friel and S. Radford, unpublished data*). For example, in Im7* the mutation L38A yields a Φ_I-value of 0.43 that decreases to 0.03 in the transition state, whilst the mutation I68V results in values for Φ_I and Φ_{TS} of 1.14 and 0.86, respectively (Figure 7). This was interpreted as evidence for stabilizing non-native interactions in the partially folded state of Im7* (*Capaldi et al., 2002*). Truncation of the equivalent residue V37 in helix II in wild-type Im9* and H2P yields a Φ_I-value that is small, ~0.2, but significantly greater than zero. Generally the Φ-values for mutations in helix II are rather small, especially in the C-terminal part of the helix which makes a great proportion of its tertiary contacts with residues from helix III (*Friel et al., 2003*). Helix II is thus at least partially formed in the intermediate observable in the folding of both wild-type Im9* and H2P. Interestingly, however, whilst the Φ_{TS}-value for this residue in Im7* folding decreases significantly, the Φ_{TS}-values obtained for the equivalent mutation in wild-type Im9* and H2P remain largely invariant, suggesting that reorganization of the partially folded state in these proteins is not so apparent. These data are substantiated by analysis of the mutation I67V in both proteins; again the Φ_I- and Φ_{TS}-values remaining largely invariant (Table 3 and Figure 7). Taken together, the data indicate, therefore, that the intermediate populated during the folding of Im9* and H2P is a three-helix bundle consisting of helices I, II and IV. If non-native interactions occur in the intermediate of these proteins they must differ from those detected in Im7*.

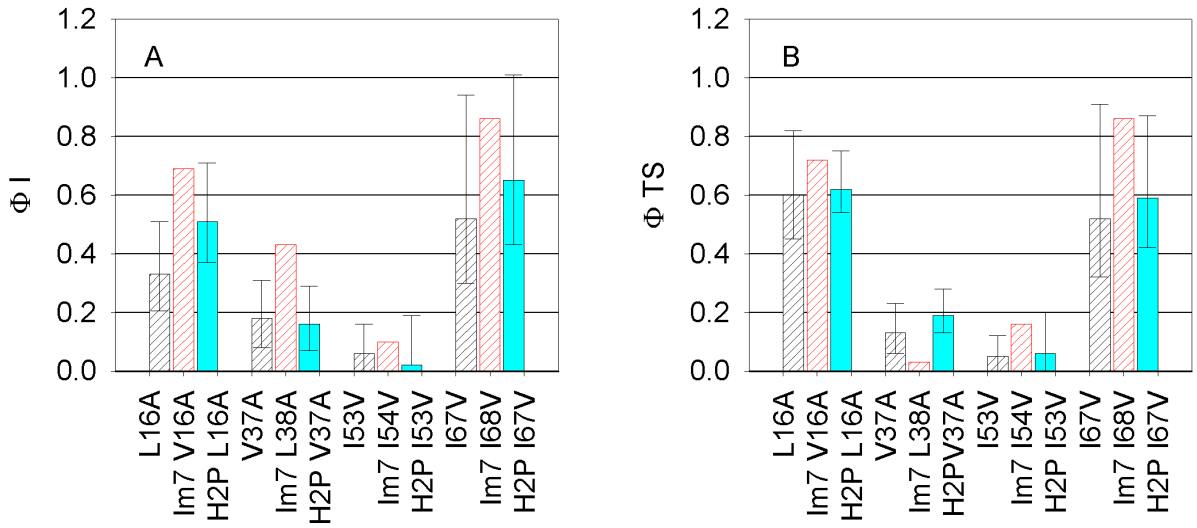


Figure 7: A) Comparison of Φ_I values of wild-type Im9* (black hatched), wild-type Im7* (red hatched) and the hydrophobicity mutant H2P (blue). **B)** Comparison of Φ_{TS} values of wild-type Im9*, wild-type Im7* and H2P. The bars are colored as in A. The Φ-values for Im7* were obtained at pH 7.0 and 0.4 M Na_2SO_4, 10°C(*Capaldi et al., 2002*). All other Φ-values were obtained at pH6.0 and 0M Na_2SO_4 at10°C. Error bars were determined as described in the Methods. Note that for all of the H2P variants studied, the fluorescence signal of the intermediate is greater than the unfolded state. In all cases the value for the signal of the intermediate ranged from 1.16 to 1.26 (normalized to the fluorescence signal of the denatured state in 8M urea (see Methods). However, large errors are introduced into the determination of the fluorescence signal of the intermediate by the destabilizing effect of the Φ-mutations.

Discussion

Im7 and Im9 are homologous bacterial proteins that have a similar four-helical structure and show approximately 60% sequence identity. Despite their similarity in structure and sequence, Im7 populates an intermediate during refolding under all conditions investigated to date (low urea concentrations at pH 5.0-7.0, both in the presence and absence of Na_2SO_4 (*Ferguson et al., 1999; Gorski et al., 2001*)), whilst an intermediate becomes populated in Im9* folding only under acidic conditions (*Ferguson et al., 1999; Gorski et al., 2001*) and data presented above).

Extensive work on Im7 has shown that the intermediate populated during refolding is on-pathway and consists of helices I, II and IV docked around a specific hydrophobic core that is stabilized by both native and non-native interactions (*Capaldi et al., 2002; Capaldi et al., 2001*). Whilst the precise nature of the non-native stabilizing interactions is not known, results from both hydrogen exchange (*Gorski et al., 2004*) and Φ-value analysis (3 of the 3 residues for which Φ_I exceeds Φ_{TS} in this protein involving residues in this helix) (*Capaldi et al., 2002*) suggest that residues in helix II play a major role in it. Here we show that the intermediate populated during the folding of Im9* at pH 6.0 resembles the intermediate formed during Im7* folding in that it is also a three-helical species containing helices I, II and IV. These species differ, however, in their fluorescence properties (the intermediate formed during Im9* folding is not hyper-fluorescent) and in their β_I values (0.82 and 0.73, respectively). In addition, the contribution of non-native interactions to the stability of these species also differs (of the two Φ-mutations that report on non-native interactions studied here neither shows a value for Φ_I that exceeds Φ_{TS}, as is observed for the equivalent residues in Im7* (*Capaldi et al., 2002*)). These data suggest that folding *via* a three-helical species is a generic feature of immunity protein folding, albeit that for two of the three family members studied to date (Im2 and Im9) the intermediate is too unstable to detect at neutral pH (*Ferguson et al., 1999; Ferguson et al., 2001*). Interestingly, an intermediate has also been detected during the folding of a chimeric immunity protein that comprises residues 1-30 and 64-86 of Im2 and residues 30-64 of Im9 (*Ferguson et al., 2001*). In addition, a second Im2/Im9 chimera shows a clear non-linearity in the logarithm of the unfolding kinetics *versus* urea concentration, a feature that is also indicative of the formation of a high-energy, on-pathway intermediate (*Bachmann and Kiefhaber, 2001; Ferguson et al., 2001*). Finally, in recent work we have shown that an intermediate in Im9* folding is stabilized at neutral pH by increasing the hydrophobicity of key core residues that stabilize this species through the formation of specific, non-native hydrophobic interactions (in the so-called 'Switch Im9' (*C. Friel and S. Radford, unpublished data*)). Now we show that the intermediate in Im9* folding can also be stabilized by increasing the helix propensity of any of the three long helices formed early during folding.

In contrast to the strategy of using rational redesign which has been used successfully to stabilize the intermediate in Im9* folding relative to both the unfolded and native states ($\Delta\Delta G_{UI}$ = 10.2 kJ mol^{-1}, $\Delta\Delta G_{UN}$ = 2.2 kJ mol^{-1} for Switch Im9* at pH 6.0 and 0.4 M Na_2SO_4 (Figure 8) (*C. Friel and S. Radford, unpublished data*), increasing helix propensity stabilizes the

intermediate and native states of Im9* to similar extents, irrespective of which helix is engineered (discounting H4 in which the native state is specifically destabilized by the removal of hydrophobic contacts involving the γCH_3 of T70) (Figure 8). Importantly, however, whilst the introduction of specific hydrophobic contacts in Switch Im9* results in substantial stabilization of the intermediate ($\Delta\Delta G_{UI}$ = 10.2 kJ mol^{-1}), altering helical propensity is much less effective ($\Delta\Delta G_{UI}$ for H1, H2 and H12 = 0.4 to -4.9 kJ mol^{-1}). Detailed interpretation of the results of the helix redesign in terms of the energetic stabilizations of individual species, however, is complex since altering helix propensity can cause ground state effects stabilizing helical structure in the unfolded state. Further work will be needed to quantify the extent to which this occurs for the mutants created here, for example a study of peptide fragments (*Munoz et al., 1996; Taddei et al., 2000; Villegas et al., 1995*) or mutants unable to fold at neutral pH , as has been performed successfully for other helical proteins. Nonetheless, the data presented here show that alterations in helical propensity stabilize the intermediate such that it becomes significantly populated during Im9* folding. This result supports the diffusion-collision model in which intermediates are predicted to be stabilized by increased helix propensity and by increase of stabilizing contacts between helices, be they either native or non-native (*Beck et al., 2001; Islam et al., 2002*). Interestingly, a similar conclusion was reached recently by a study on a series of three-helical proteins varying in their helical propensity. The folding properties of topologically similar three-helix bundle proteins from the homeodomain-like protein family with different helical stabilities (as calculated by the AGADIR program (*Munoz and Serrano, 1997*)) were investigated. It was shown that single-site mutants that increase helical propensities of proteins exhibiting low helical propensity increases the population of the intermediate such that it becomes experimentally visible (*Gianni et al., 2003*). Careful analysis can determine whether secondary and tertiary structure form concomitantly or sequentially in the studied proteins. It was demonstrated that in the protein with higher helical propensity, secondary structure forms first and consolidation of packing interactions and water expulsion occurs only after the transition state. This is in strong support of a unified model of folding, in which folding is either more or less hierarchical, depending on the stability of secondary structure (*Daggett and Fersht, 2003*).

Another design strategy to create an Im9* species that folds with three-state kinetics employed in the presented study was to increase the hydrophobicity of the solvent exposed residues in helix II (Figure 1B, C, 2C, F,). The most striking observation was that increasing the

hydrophobicity of solvent exposed residues in helix II by the mutations K35L and E42L stabilizes both the folding intermediate and the native state (Figure 8). This by contrast is not observed when the hydrophobicity of solvent exposed residues in helices I and IV is increased, suggesting that helix II plays a unique role in the folding of Im9*. The effect of increasing the hydrophobicity of solvent exposed residues thus depends critically on the location of the mutations. Sequence changes that increase the hydrophobicity of solvent exposed residues in helices I and IV merely cause a negative hydrophobic effect by stabilizing the unfolded state. Increasing the hydrophobicity of residues in the centre of helix II however, stabilizes the intermediate, as well as the rate-limiting transition states and the native protein (Figure 8). This demonstrates that the differences in the kinetic folding mechanisms of Im7* and Im9* cannot be rationalized simply by differences in the hydrophobicity of their sequences. In addition, it shows that the stability of the intermediate populated during the folding of Im9* (and by analogy Im7*) cannot be attributed to random collapse of hydrophobic residues early in folding, in accord with previous results which showed that truncation of solvent exposed residues in Im7* has no destabilizing effect on the intermediate or native state ($\Delta\Delta G_{UI}$ = -1.2-0 kJ mol^{-1}, $\Delta\Delta G_{UN}$ = -2.1-1 kJ mol^{-1} (*Capaldi et al., 2002*)). Interestingly, Switch Im9* involves the mutation of three residues, V37 (helix II) and V71 (helix IV) which are buried in the core of native Im9*, and E41, a solvent exposed residue in helix II (*C. Friel and S. Radford, unpublished data*). The latter was included in Switch Im9*, since its mutation to Leu increases the fluorescence signal of the intermediate (such that it becomes hyper-fluorescent). It also stabilizes the intermediate, presumably by decreasing unfavorable electrostatic interactions or by allowing helix II to optimize the burial of hydrophobic surface area in the developing hydrophobic core (*C. Friel and S. Radford, unpublished data*). This mutation also stabilizes the native protein by 3.7 kJ mol^{-1} (at pH 6.0, 0.4 M Na_2SO_4 and 10 ºC (*C. Friel and S. Radford, unpublished data*)) providing a platform to explain the folding properties of the variant H2P. In H2P the intermediate is stabilized by 7.7 kJ mol^{-1} at pH 6.0 and becomes hyper-fluorescent while the native state is stabilized by 3.6 kJ mol^{-1} as well (Figure 8)). These data suggest, therefore, that the switch in folding kinetics from two- to three-state for the variant H2P can be attributed to the mutation of E42L mirroring the effect of the substitution E41V. On one hand these data highlight the importance of helix II in determining the kinetic folding mechanism of the immunity proteins and demonstrate that even a single amino acid substitution that stabilizes the developing core through the burial of non-native hydrophobic surface area can effectively switch the folding of this protein from two- to three-state. On the other hand it highlights that solvent exposed polar residues can be important for the folding pathway of a protein. The

surface exposed polar residues in Im9 may prevent the rocking of helix II that may stabilize the intermediate in H2P and also in Im7. A similar effect was observed in a study on a designed four-helix bundle where the energy of the native state remained unchanged because the decrease in entropic penalty was compensated by an equivalent decrease in enthalpic gain. The prominent effect in this case was the stabilization of non-native folded states, a reduction of the bias gap and thus the inability of the protein to adopt a unique conformation (*Hill and DeGrado, 2000*).

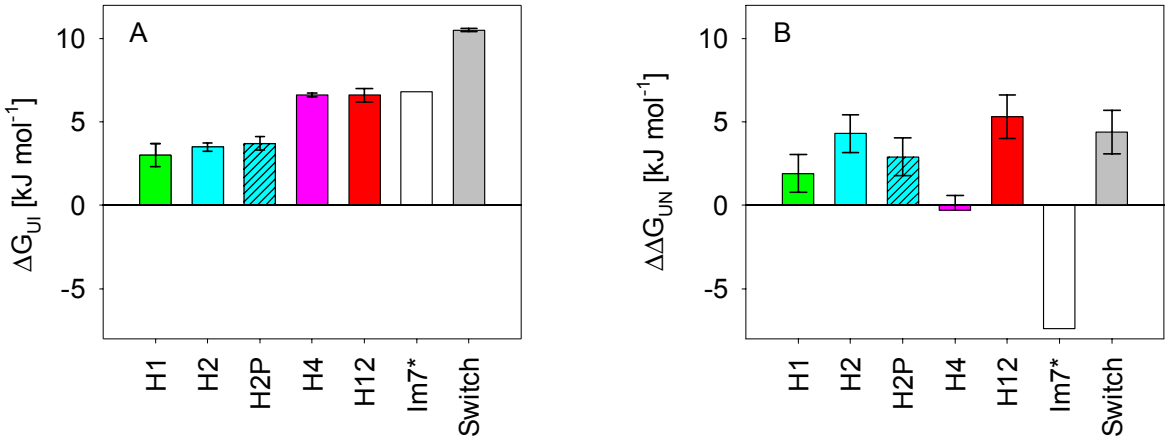


Figure 8 A) Summary of the stability of the intermediates populated during the folding of different immunity proteins. All ΔG_{UI} values are measured at pH 7.0 in the absence of Na_2SO_4(*Ferguson et al., 1999; Friel, 2004*). Note that under these conditions wild-type Im9* folds with two-state kinetics (i.e. ΔG_{UI} is less than 0.7 kJ mol^{-1}). **B**) As A, but the stabilization of the native state of each protein relative to wild-type Im9* ($\Delta\Delta G_{UN}$) is plotted.

The presented study indicates that local interactions, as the main determinants of helical structure, play an important role in determining the folding properties and the stability of helical proteins. Stabilizing any of the three long helices of Im9* stabilizes the native state and every other state in which the stabilized helix is already structured. However, when a helix becomes formed in the unfolded state as well as in all structured states there will be no observable effect on relative stability. Thus, as indicated by theoretical studies (*Abkevich et al., 1995; Doyle et al., 1997; Govindarajan and Goldstein, 1995; Unger and Moult, 1996*), it is possible but problematic to stabilize elements of secondary structure in order to stabilize the native state relative to the unfolded state. The question is why some helices in Im9* derivatives become structured in the unfolded state and others do not. After the mutational stabilization the central residues of helix II show a helical propensity of nearly 30% (Figure 2B) but the data

does not indicate that the designed helix becomes structured in the unfolded state. The stabilized helix I in contrast never shows a higher helical propensity than 20% and still the data clearly indicates that helix I is structured in the unfolded state of the mutant H1. Results obtained with the mutant H12 indicate that the stability of a helix in the unfolded state depends not only on the energetic contribution of the local interactions but also on non-local interactions that can develop even in this early stage of folding. This confirms the importance of the structure of an unfolded state of a protein for its folding pathway and suggest that a more thorough characterization of the unfolded state of Im9* and its various derivatives is probably worthwhile.

pH6	K_{ui}[a]	M_{ui}[b]	β_I[c]	k_{in}	m_{in}	β_{TS}[d]	k_{ni}	m_{ni}	□ G_{ui}[e]	□ G_{un}[f]	M_{un}[g]
Im9	3.8±0.6	4.5±0.15	0.82±0.04	440±50	0.6±0.1	0.93±0.05	0.12±.001	-0.38±0.04	-3.1±0.4	-22.4±0.9	5.48±0.18
H1	4.4±0.3	4.2±0.1	0.83±0.03	480±20	0.5±0.1	0.93±0.04	0.185±0.01	-0.38±0.02	-3.5±0.18	-22.0±0.4	5.08±0.14
H2	12.5±0.8	4.3±0.15	0.81±0.04	500±50	0.48±0.03	0.91±0.04	0.1±0.004	-0.5±0.03	-5.9±0.16	-26.0±0.5	5.28±0.16
H4	29.4±5	4.3±0.1	0.83±0.03	290±20	0.5±0.07	0.92±0.03	1.42±0.1	-0.4±0.03	-8.0±0.44	-20.5±0.8	5.20±0.13
H12	6.3±0.4	3.6±0.2	0.80±0.06	650±100	0.5±0.1	0.92±0.07	0.2±0.02	-0.38±0.06	-4.3±0.16	-23.3±0.8	4.48±0.23
pH7											
Im9[h]	-	-		1200±100	4.9±0.05	0.92±0.02	0.012±0.002	-0.42±0.1	-	-27.1±0.8	5.3±0.11
H1	3.6±0.8	4.2±0.1	0.83±0.03	1000±100	0.45±0.1	0.92±0.04	0.016±0.002	-0.42±0.1	-3.0±0.68	-29.0±0.8	5.07±0.17
H2	4.44±0.9	4.1±0.1	0.84±0.03	1100±100	0.38±0.05	0.91±0.04	0.008±0.001	-0.42±0.1	-3.5±0.25	-31.4±0.8	4.9±0.15
H4	16.7±2.4	4.2±0.1	0.83±0.03	550±50	0.45±0.08	0.92±0.04	0.105±0.001	-0.42±0.2	-6.6±0.12	-26.8±0.4	5.1±0.17
H12	16.7±2.4	4.1±0.7	0.85±0.03	980±100	0.3±0.05	0.91±0.04	0.017±0.003	-0.42±0.1	-6.6±0.4	-32.4±1.0	4.8±0.13

[a] $K_{UI} = k_{UI}/k_{IU}$. [b] $M_{UI} = m_{UI}+m_{IU}$. [c] $\beta_I = M_{UI}/M_{UN}$. [d] $\beta_{TS} = (M_{UI} + m_{IN}) /M_{UN}$. [e] $\Delta G_{UI} = -RT \ln(K_{UI})$. [f] $\Delta G_{UN} = -RT \ln(K_{UI} \times (k_{IN}/k_{NI}))$. [g] $M_{UN} = M_{UI} + m_{IN}-m_{NI}$. [h] The data for Im9 WT at pH7.0 was fitted to a model describing a two-state transition where k_{IN} (m_{NI}) and k_{NI} (m_{NI}) are k_{UI} (m_{UI}) and k_{IU} (m_{IU}), respectively, and $\Delta G_{UN} = -RT \ln(k_{UN}/k_{NU})$. Rates are in s^{-1}, stabilities are in kJ mol^{-1}, m-values are in kJ $mol^{-1}M^{-1}$.

Table I: Folding and unfolding parameters determined for wild-type Im9* and the variants H1, H2, H4 and H12 created in this study. The kinetics were measured in buffer A (see Methods) at 10ºC. The data were fitted to a model describing a three-state folding transition, except for wild-type Im9* at pH7.0 which was fitted to a two-state model.

pH6	K_{ui}[a]	M_{ui}[b]	β_I[c]	k_{in}	m_{in}	β_{TS}[d]	k_{ni}	m_{ni}	ΔG_{ui}[e]	ΔG_{un}[f]	M_{un}[g]
Im9	3.8±0.6	4.5±0.15	0.82±0.04	440±50	0.6±0.1	0.93±0.05	0.12±.001	-0.38±0.04	-3.1±0.4	-22.4±0.9	5.48±0.18
H1P[e]				450±70	4.9±0.2	0.95±0.05	0.164±0.015	-0.27±0.03		-18.6±0.6	5.17±0.2
H2P	26.5±4.1	4.76±0.1	0.80±0.03	190±20	0.9±0.15	0.95±0.04	0.082±0.005	-0.3±0.03	-7.7±0.4	-26.0±0.8	5.96±0.18
H4P[e]				190±40	5.2±0.2	0.93±0.05	0.15±0.02	-0.38±0.06		-16.8±0.9	5.6±0.21
pH7											
Im9[h]	-	-		1200±100	4.9±0.05	0.92±0.02	0.012±0.002	-0.4±0.1	-	-27.1±0.8	5.3±0.11
H1P[e]				530±50	5±0.15	0.94±0.04	0.018±0.002	-0.32±0.03		-24.2±0.5	5.32±0.15
H2P	4.8±0.8	4.52±0.1	0.83±0.03	520±50	0.48±0.08	0.92±0.04	0.0074±0.0005	-0.42±0.1	-3.7±0.4	-30.0±0.8	5.42±0.16
H4P[e]				200±30	5±0.15	0.92±0.04	0.0085±0.0015	-0.46±0.04		-23.7±0.8	5.46±0.16

[a] $K_{UI} = k_{UI}/k_{IU}$. [b] $M_{UI} = m_{UI}+m_{IU}$. [c] $\beta_I = M_{UI}/M_{UN}$. [d] $\beta_{TS} = (M_{UI} + m_{IN}) /M_{UN}$. [e] $\Delta G_{UI} = -RT \ln(K_{UI})$. [f] $\Delta G_{UN} = -RT \ln(K_{UI} \times (k_{IN}/k_{NI}))$. [g] $M_{UN} = M_{UI} + m_{IN}-m_{NI}$. [h] The data for the respective mutants was fitted to a model describing a two-state transition where k_{IN} (m_{NI}) and k_{NI} (m_{NI}) are k_{UI} (m_{UI}) and k_{IU} (m_{IU}), respectively, and $\Delta G_{UN} = -RT \ln(k_{UN}/k_{NU})$. Rates are in s^{-1}, stabilities are in kJ mol^{-1}, m-values are in kJ $mol^{-1}M^{-1}$.

Table II: Folding and unfolding parameters determined for wild-type Im9* and the variants H1P, H2P and H4P created in this study. The kinetics were measured in buffer A (see Methods) at 10ºC. The data were fitted to a model describing a two-state folding transition, except for wild-type Im9* at pH6.0 and H2P at pH6.0 and 7.0 which were fitted to a three-state on-pathway model (see methods).

Mutant	Position[a]	K_{ui}[b] (M_{ui}[c])	k_{in} (m_{in})	k_{ni} (m_{ni})	ΔG_{ui}[d]	ΔG_{un}[e]	Φ_I[f]	Φ_{TS}[g]
Im9*		3.8 (4.5)	440 (0.6)	0.12 (-0.38)	-3.1 ± 0.4	-22.4 ±0.9		
L16A	Helix I (p)	1.3 ± 0.1 (4.8 ± 0.1)	190 ± 15 (0.5 ± 0.15)	0.42 ± 0.02 (-0.43 ± 0.02)	-0.7 ± 0.3	-15.06 ± 0.6	0.33+0.18 -0.125	0.60+0.22 -0.15
V37A	Helix II (b)	1.8 ± 0.3 (4.5 ± 0.2)	530±50 (0.9 ± 0.2)	4 ± 0.3 (-0.05 ±0.03)	-1.4 ± 0.4	-12.9 ± 0.8	0.18+0.13 -0.1	0.13 +0.1 -0.07
I53V	Helix III (b)	3.1 ± (0.3) (4.9 ± 0.2)	450 ± 30 (0.4 ±0.1)	4 ± 0.2 (-0.15± 0.02)	-2.6 ± 0.3	-13.8 ± 0.5	0.06+0.1 -0.07	0.05+0.07 -0.05
I67V	Helix IV (b)	1.3 ± 0.1 (4.5 ± 0.2	440 ± 40 (0.7 ± 0.2	0.32 ± 0.02 (-0.51 ± 0.03	-0.7 ± 0.3	-17.7 ± 0.6	0.52+0.42 -0.22	0.52+0.39 -0.20
H2P		26.5 (4.76)	190 (0.9)	0.082 (-0.3)	-7.1	-25.3		
H2P L16A	Helix I (p)	3.33 ± 0.5 (4.65± 0.2)	125 ± 10 (0.95 ± 0.15)	0.39 ± 0.03 (-0.33 ± 0.03)	-2.8 ± 0.4	-16.4 ± 0.8	0.51+0.2 -0.14	0.62+0.13 -0.08
H2P V37A	Helix II (b)	14.3 ± 2.1 (4.7 ± 0.1)	170 ± 10 (0.9 ± 0.1)	2 ± 0.1 (-0.15 ± 0.05)	-6.3 ± 0.4	-16.7 ± 0.6	0.16+0.13 -0.09	0.19+0.09 -0.06
H2P I53V	Helix III (b)	18.2 ± 2.7 (4.6 ± 0.1)	180 ± 20 (0.95 ± 0.1)	1 ± 0.1 (-0.39 ± 0.03)	-6.8 ± 0.4	-19.0 ± 0.9	0.02+0.17 -0.1	0.06+0.14 -0.08
H2P I67V	Helix IV (b)	4.1 ± 0.6 (4.6 ± 0.1)	230 ± 20 (0.85 ± 0.15)	0.27 ± 0.03 (-0.4 ± 0.04)	-3.3 ± 0.4	-19.2 ± 0.8	0.65+0.36 -0.22	0.59 + 0.28 - 0.17

[a] (b) buried; (p) partially buried. [b]$K_{UI} = k_{UI}/k_{IU}$. [c]$M_{UI} = m_{UI}+m_{IU}$. [d]$\Delta G_{UI} = -RT \ln(K_{UI})$. [e]$\Delta G_{UN} = -RT \ln(K_{UI} \times (k_{IN}/k_{NI}))$. [f]$\Phi_I = (\Delta G_{ui}(\text{mutant})-(\Delta G_{ui}(\text{wt}))/ (\Delta G_{un}(\text{mutant})-(\Delta G_{un}(\text{wt}))$. [g]$\Phi_{TS} = ((\Delta G_{ui}(\text{mutant}) - (\Delta G_{ui}(\text{wt}))) - (RT \ln (k_{in}(\text{mutant}) / k_{in}(\text{wt}))) / (\Delta G_{un}(\text{mutant})-(\Delta G_{un}(\text{wt}))$. Rates are in s^{-1}, stabilities are in kJ mol^{-1}, m-values are in kJ $mol^{-1}M^{-1}$.

Table III: Parameters determined from the best fit of the folding/unfolding kinetics for wild-type Im9*, H2P and their variants at pH6.0, 10ºC to a three-state model (see Methods). Values for Φ_I and Φ_{TS} were calculated in 0M urea.

References

Abkevich V.I., Gutin A.M. and Shakhnovich E.I. (1995) J Mol Biol, 252: 460-471

Bachmann A. and Kiefhaber T. (2001) J Mol Biol, 306: 375-386

Bai Y. (1999) J Biomol NMR, 15: 65-70

Baldwin R.L. and Rose G.D. (1999i) Trends Biochem Sci, 24: 26-33

Baldwin R.L. and Rose G.D. (1999ii) Trends Biochem Sci, 24: 77-83

Beck C., Siemens X. and Weaver D.L. (2001) Biophys J, 81: 3105-3115

Bennion B.J. and Daggett V. (2003) Proc Natl Acad Sci U S A, 100: 5142-5147

Capaldi A.P., Kleanthous C. and Radford S.E. (2002) Nat Struct Biol, 9: 209-216

Capaldi A.P., Shastry M.C., Kleanthous C., Roder H. and Radford S.E. (2001) Nat Struct Biol, 8: 68-72

Chowdhury S., Zhang W., Wu C., Xiong G. and Duan Y. (2003) Biopolymers, 68: 63-75

Daggett V. and Fersht A.R. (2003) Trends Biochem Sci, 28: 18-25

Dennis C.A., Videler H., Pauptit R.A., Wallis R., James R., Moore G.R. and Kleanthous C. (1998) Biochem J, 333 (Pt 1): 183-191

Dill K.A. (1990) Biochemistry, 29: 7133-7155

Doyle R., Simons K., Qian H. and Baker D. (1997) Proteins, 29: 282-291

Ferguson N., Capaldi A.P., James R., Kleanthous C. and Radford S.E. (1999) J Mol Biol, 286: 1597-1608

Ferguson N., Li W., Capaldi A.P., Kleanthous C. and Radford S.E. (2001) J Mol Biol, 307: 393-405

Fersht A.R., Matouschek A. and Serrano L. (1992) J Mol Biol, 224: 771-782

Friel C.T. (2004):

Friel C.T., Capaldi A.P. and Radford S.E. (2003) J Mol Biol, 326: 293-305

Gianni S., Guydosh N.R., Khan F., Caldas T.D., Mayor U., White G.W., DeMarco M.L., Daggett V. and Fersht A.R. (2003) Proc Natl Acad Sci U S A, 100: 13286-13291

Gill S.C. and von Hippel P.H. (1989) Anal Biochem, 182: 319-326

Gong H., Isom D.G., Srinivasan R. and Rose G.D. (2003) J Mol Biol, 327: 1149-1154

Gorski S.A., Capaldi A.P., Kleanthous C. and Radford S.E. (2001) J Mol Biol, 312: 849-863

Gorski S.A., Le Duff C.S., Capaldi A.P., Kalverda A.P., Beddard G.S., Moore G.R. and Radford S.E. (2004) J Mol Biol, 337: 183-193

Govindarajan S. and Goldstein R.A. (1995) Proteins, 22: 413-418

Hill R.B. and DeGrado W.F. (2000) Structure Fold Des, 8: 471-479

Islam S.A., Karplus M. and Weaver D.L. (2002) J Mol Biol, 318: 199-215
Kraulis P.J. (1991) J Appl Crystallog, 24: 946-950
Lacroix E., Viguera A.R. and Serrano L. (1998) J Mol Biol, 284: 173-191
Lopez-Hernandez E., Cronet P., Serrano L. and Munoz V. (1997) J Mol Biol, 266: 610-620
Merrit E.A., Bacon, D.J. (1997) Methods Enzymol, 277: 505-524
Munoz V., Cronet P., Lopez-Hernandez E. and Serrano L. (1996) Fold Des, 1: 167-178
Munoz V., Lopez E.M., Jager M. and Serrano L. (1994) Biochemistry, 33: 5858-5866
Munoz V. and Serrano L. (1994) Nat Struct Biol, 1: 399-409
Munoz V. and Serrano L. (1995) J Mol Biol, 245: 275-296
Munoz V. and Serrano L. (1997) Biopolymers, 41: 495-509
Northey J.G., Di Nardo A.A. and Davidson A.R. (2002) Nat Struct Biol, 9: 126-130
Osborne M.J., Breeze A.L., Lian L.Y., Reilly A., James R., Kleanthous C. and Moore G.R. (1996) Biochemistry, 35: 9505-9512
Otzen D.E., Itzhaki L.S., elMasry N.F., Jackson S.E. and Fersht A.R. (1994) Proceedings Of The National Academy Of Sciences Of The United States Of America, 91: 10422-10425
Pakula A.A. and Sauer R.T. (1990) Nature, 344: 363-364
Poso D., Sessions R.B., Lorch M. and Clarke A.R. (2000) J Biol Chem, 275: 35723-35726
Richards F.M. and Lim W.A. (1993) Q Rev Biophys, 26: 423-498
Rose G.D. and Roy S. (1980) Proc Natl Acad Sci U S A, 77: 4643-4647
Taddei N., Chiti F., Fiaschi T., Bucciantini M., Capanni C., Stefani M., Serrano L., Dobson C.M. and Ramponi G. (2000) J Mol Biol, 300: 633-647
Takei J., Pei W., Vu D. and Bai Y. (2002) Biochemistry, 41: 12308-12312
Unger R. and Moult J. (1996) J Mol Biol, 259: 988-994
Uversky V.N. (2002) FEBS Lett, 514: 181-183
Viguera A.R., Vega C. and Serrano L. (2002) Proc Natl Acad Sci U S A, 99: 5349-5354
Viguera A.R., Villegas V., Aviles F.X. and Serrano L. (1997) Fold Des, 2: 23-33
Villegas V., Viguera A.R., Aviles F.X. and Serrano L. (1996) Fold Des, 1: 29-34
Villegas V.V., Viguera A.R., Aviles F.X. and Serrano L. (1995) Fold Des, 1: 29-34

3.1. Appendix to "Helix stability and Hydrophobicity in the Folding Mechanism of the Bacterial Immunity Protein Im9."

The manuscript presented on the previous pages describes various mutagenesis strategies for stabilizing an intermediate in the folding of Im9*. One strategy is to decrease the polarity of the surface of helix II. As described in the manuscript several Φ-mutations were introduced to test whether the intermediate that appears upon increasing hydrophobicity of helix II is similar to the intermediate seen in the folding of Im9* at lower pH. We could confirm that both intermediates take the form of a three-helix bundle, in which the non-native interactions seen in the related proteins Im7 and Switch could not be detected.

Another way of stabilizing an intermediate in Im9* is by stabilizing its native helical structure. It has been shown that the intermediate that occurs in the folding of the homologous Im7 protein is a three-helix bundle consisting of helices I, II and IV and that three of the four helices in Im9* are structured already in the transition state. Stabilizing helices in Im9* might thus stabilize a previously hidden intermediate and make it observable. However, the native and transition states and in fact all states containing structured helices will be similarly stabilized by this approach.

Stabilizing helix IV in Im9* proved to be difficult because many of the solvent exposed positions are occupied by residues that show a high preference for helical Φ– and Ψ–angles. The capping residues are proline and glycine, the former is probably important for specifying the extent of the helix while changing glycine is known to cause strong ground state effects. Eventually the mutation T70A was included in the design. This is predicted to increase the stability of helix IV by 10% but deletes a hydrophobic contact that is important for the native state. Accordingly, the native state of the resulting mutant H4 is destabilized even as the intermediate is significantly stabilized. As in the mutant H2P, several Φ-mutations were introduced in H4 to confirm that the observed intermediate is the expected three-helix bundle. Due to restrictions in length, this data is not included in the manuscript.

The Φ-values measured with H4 show that the intermediate shows a slightly different structure than expected (Table 3.1., Figure 3.1 A). The Φ_I-value for the mutation L16A is virtually 0 and thus much smaller than the Φ-value for this mutation in WT Im9*, H2P or Im7. This does not necessarily mean that helix I does not participate in the intermediate. Helix I could be folded in the H4 intermediate but oriented differently compared to what has been observed for other intermediates. To test this we introduced the mutation A13G that probes the structure of the backbone of helix I in the mutant H4. The high Φ-value obtained for A13G indicates that helix I is indeed structured in the H4 intermediate. Because the helical stability of helix I is low, this could not be the case if it were not stabilized by non-local interactions with the rest of the intermediate. This suggests that in H4, helix I is docked to rest of the structure in a different orientation. The Φ_I-value for V37A is very low indicating that non-native contacts as observed in Im7* folding do not occur in this intermediate. The Φ_I-value of 0.03 observed for I53V indicates that helix III is not structured in this intermediate. For the mutation I67V we measure a Φ_I-value very similar to the Φ_I-value found for Im9* WT indicating that helix IV is as well structured in the H4 intermediate as it is in the WT Im9* intermediate. The role of helix I in the H4 intermediate might be understood by a more thorough Φ-value analysis in H4. This could explain how helix I interacts with the rest of the intermediate and reveal if other tertiary contacts are altered in H4.

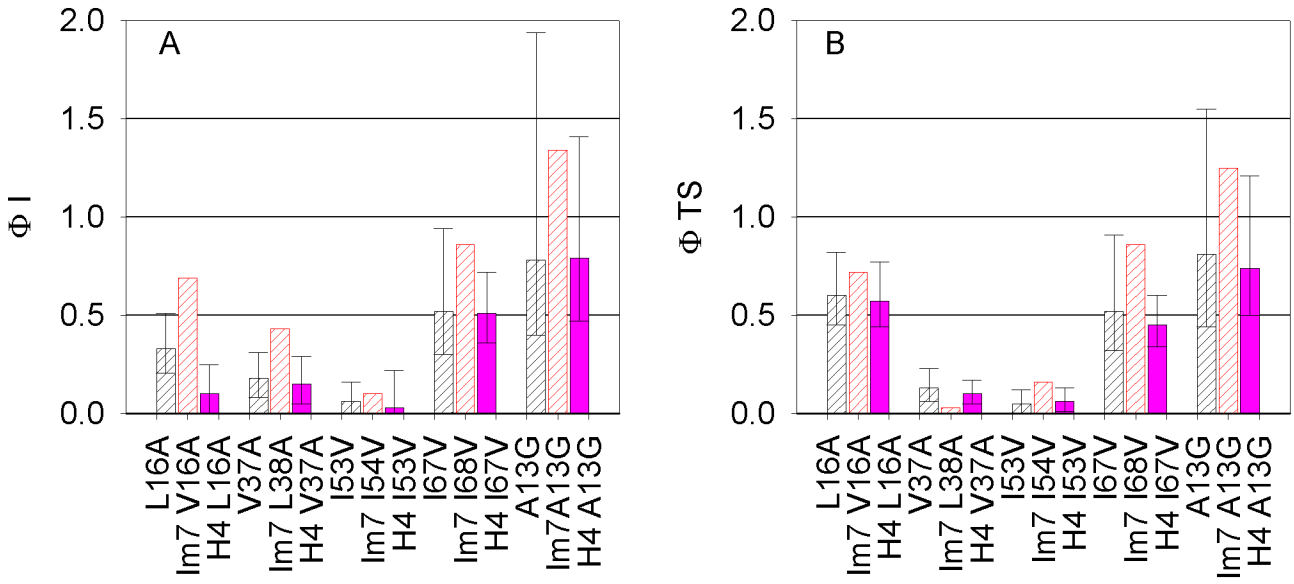


Figure 3.1 A) Comparison of Φ_I values of WT Im9* (black hatched), WT Im7* (red hatched) and the helicity mutant H4 (pink).B) Comparison of Φ_{TS} values of WT Im9*, WT Im7* and the helicity mutant H4 (pink). The bars are coloured as in A. All Φ-values for Im7* were obtained at pH7.0 and 0.4 M Na_2SO_4, 10 °C (*Capaldi et al., 2002*), all other Φ-values were obtained at pH 6.0 and 0 M Na_2SO_4. Error bars were determined as described in the Methods.

Apart from the Φ_I-values discussed above, we also determined the Φ-values for transition state II in H4 (Figure 1B). Here the data obtained for H4 resembles the values that were obtained with Im9* at pH 6.0. Thus H4 folds through the same transition state Im9* WT and H2P in spite of having a slightly different intermediate.

Mutant	Position[a]	K_{ui}[b] (M_{ui}[c])	k_{in} (m_{in})	k_{ni} (m_{ni})	ΔG_{ui}[d]	ΔG_{un}[e]	Φ_I[f]	Φ_{TS}[g]
Im9*		3.8 (4.5)	440 (0.6)	0.12 (-0.38)	-3.1 ± 0.4	-22.4 ±0.9		
L16A	Helix I (p)	1.3 ± 0.1 (4.8 ± 0.1)	190 ± 15 (0.5 ± 0.15)	0.42 ± 0.02 (-0.43 ± 0.02)	-0.7 ± 0.3	-15.06 ± 0.6	0.33+0.18 -0.125	0.60+0.22 -0.15
V37A	Helix II (b)	1.8 ± 0.3 (4.5 ± 0.2)	530±50 (0.9 ± 0.2)	4 ± 0.3 (-0.05 ±0.03)	-1.4 ± 0.4	-12.9 ± 0.8	0.18+0.13 -0.1	0.13 +0.1 -0.07
I53V	Helix III (b)	3.1 ± (0.3) (4.9 ± 0.2)	450 ± 30 (0.4 ±0.1)	4 ± 0.2 (-0.15± 0.02)	-2.6 ± 0.3	-13.8 ± 0.5	0.06+0.1 -0.07	0.05+0.07 -0.05
I67V	Helix IV (b)	1.3 ± 0.1 (4.5 ± 0.2	440 ± 40 (0.7 ± 0.2	0.32 ± 0.02 (-0.51 ± 0.03	-0.7 ± 0.3	-17.7 ± 0.6	0.52+0.42 -0.22	0.52+0.39 -0.20
A13G	Helix 1 (e)	1.3 ± 0.3 (4.5 ± 0.15)	420 ± 40 (0.7± 0.2)	0.155 ± 0.01 (-0.38 ± 0.04)	-0.7 ±0.3	-19.3 ± 0.6	0.78 +1.16 -0.38	0.81+1.03 -0.37
H4		29.4 (4.3)	290 (0.5)	1.42 (-0.4)	-8.0	-20.5		
H4 L16A	Helix I (p)	22.4 ± 2.4 (4.6 ± 0.2)	72 ± 5 (0.6 ± 0.15)	4.9 ± 0.2 (-0.26 ± 0.02)	-7.4 ± 0.2	-13.6 ± 0.5	0.1+0.15 -0.10	0.57+0.2 -0.13
H4 V37A	Helix II (b)	16.7 ± 2.5 (4.6 ±0.2)	350 ± 30 (0.8 ±0.2)	42 ± 3 (0.05±0.03)	-6.6 ± 0.4	-11.6 ± 0.8	0.15 +0.14 -0.10	0.10+0.07 -0.05
H4 I53V	Helix III (b)	26.7 ± 4.0 (4.3 ± 0.2)	265 ± 15 (0.5 ± 0.1)	23.5 ± 2 (-0.2± 0.03)	-7.7 ± 0.4	-13.4 ± 0.7	0.03+0.16 -0.10	0.06 +0.07 -0.05
H4 I67V	Helix IV (b)	6.9 ± 0.4 (4.3 ± 0.1)	340 ± 20 (0.5 ± 0.05)	6.8 ± 0.4 (-0.15 ± 0.02)	-4.5 ± 0.1	-13.7 ± 0.4	0.51+0.21 -0.15	0.45+0.15 -0.11
H4 A13G	Helix 1 (e)	7.3 ± 0.7 (4.3 ± 0.1)	320 ± 20 (0.45 ± 0.1)	2.25 ± 0.2 (-0.2 ± 0.08)	-4.7 ± 0.2	-16.4 ± 0.6	0.79+0.62 -0.32	0.74+0.47 -0.24

[a] (b) buried; (p) partially buried; (e) exposed. [b] $K_{UI} = k_{UI}/k_{IU}$. [c] $M_{UI} = m_{UI}+m_{IU}$. [d] $\Delta G_{UI} = -RT \ln(K_{UI})$. [e] $\Delta G_{UN} = -RT \ln(K_{UI} \times (k_{IN}/k_{NI}))$. [f] $\Phi_I = (\Delta G_{ui}(\text{mutant}) - (\Delta G_{ui}(\text{wt}))/ (\Delta G_{un}(\text{mutant}) - (\Delta G_{un}(\text{wt}))$. [g] $\Phi_{TS} = ((\Delta G_{ui}(\text{mutant}) - (\Delta G_{ui}(\text{wt}))) - (RT \ln (k_{in}(\text{mutant}) / k_{in}(\text{wt}))) / (\Delta G_{un}(\text{mutant}) - (\Delta G_{un}(\text{wt}))$. Rates are in s^{-1}, stabilities are in kJ mol^{-1}, m-values are in kJ $mol^{-1}M^{-1}$.

Table 3.1. Kinetic parameters of the WT Im9* and H4 and their respective Φ-mutants at pH6.0 under standard buffer conditions. Errors are estimated limits of fit. All rate constants are given in units of s^{-1}. All m-values are in units of $kJmol^{-1}M^{-1}$. All stabilities are in kJ mol^{-1}.

4. PROTEIN-DNA INTERACTIONS

The protein folding process and its analysis were described in great length in the previous chapters. The ultimate biological reason for the importance of this process is that virtually only folded proteins show function (*Pain, 1994*). One of the functions that folded proteins exhibit is specific DNA-binding activity. When proteins bind only to specific DNA sequences or particular DNA conformations then the recognition of these sequences or conformations is an integral part of the binding process (*Stryer, 1990*).

One process in which protein-DNA interactions are crucial is the replication of the genome that precedes cell division. Another process that is largely based on sequence specific interactions between protein and DNA is differential gene expression. Differential gene expression means that the cells of a metazoan organism express different proteins in spite of having the same genetic material at their disposal. This is essential for growth and development, cell cycle regulation and the differentiation of the various tissues and organs that make up an organism. During the differentiation process pluripotent cells like the merismatic cells in plants and the stem cells in animals engage in differential gene expression and so specialize to fulfill certain roles in the organism (*Alberts, 2002*).

Whether a protein encoded by a certain gene is actually present in the cytoplasm of a certain cell at a certain time, can be regulated on many different levels by many different mechanisms. Generally, the transcription of a gene, the process in which a macromolecular complex synthesizes an m-RNA transcript of a stretch of DNA, is the most important level of control. On the most basic level, transcriptional control consist of an operator, which is a DNA segment with regulatory function and a transcription factor, which can bind to the operator. Depending on the localization of the operator relative to the promoter and the properties of the transcription factor itself, binding of the transcription factor can activate/enhance or prevent/decrease transcription (*Alberts, 2002*).

Protein-DNA interactions are not only important in the direct control of transcription but also for enzymatic reactions involving DNA, i.e. site-specific methylation of the DNA backbone or site-specific nuclease activity, repair of DNA damaged by mutagenic factors and for mediating the cell-cycle specific changes that occur in DNA structure.

4.1. Biological Role of Transcription Factors

In the last years, the entire genome of a number of organisms has been sequenced. Subsequently it became clear that the sequence of the genome of an organism could not describe this organism as a functional entity. For example, comparing the genomes of organism as diverse as nematodes, mice and humans reveals a surprising similarity of genetic content. The genome of vertebrates contains only twice as many genes as that of invertebrates and even this increase is mostly due to duplication of genes. Obviously morphological and behavioral complexity is not only a function of the number of genes but also of the regulation of the genes. A good illustration of this is the fact that the yeast genomes encodes approximately 300 transcription factors and approximately 6000 genes whereas *H. sapiens* has as many as 3000 transcription factors that regulate as few as 26 000 genes (*Levine and Tjian, 2003; Venter et al., 2001*). Understanding how the genome is regulated will ultimately help to understand health problems such as cancer and abnormal fetal development (*Miletich and Sharpe, 2003; Packham and Brook, 2003*) or viral infection (*Wu and Marsh, 2003*) in which gene regulation is impaired or under control of a virus. A more thorough understanding of the role of transcription factors in these processes can lead to new approaches in treatment and prevention of viral infections, cancers and developmental problems (*Uil et al., 2003; Waddick and Uckun, 1998; Waddick and Uckun, 1999*).

4.2. Biochemistry of DNA Recognition and Binding

One way of understanding the interaction between DNA and protein are biochemical experiments. Such experiments can be conducted under clearly defined conditions, where the reactants and their respective concentration, the reaction buffer and the temperature can be strictly controlled. By such methods various biochemical parameters of the binding reaction, such as ΔH, ΔG, stoichiometry of the reaction, the rate of the reaction as a function of the reactants and the temperature and osmotic strength dependence of all of these parameters, can be measured. The drawback of these measurements is that it is difficult to gauge whether the measured parameters will have meaning under in-vivo conditions. Firstly, the high concentration of macromolecules in a cell gives rise to unspecific crowding effects. Secondly, the interaction between a certain DNA-binding protein and its target site is on one hand dependent on the conformation of the DNA (A-, B-, Z-DNA) and its packing status, on the other hand the affinity of the protein to its target site may be modulated in a specific manner by other proteins present in the nucleus (*Alberts, 2002*). Another hurdle for site-specific

binding of protein to DNA that is often not taken into account in biochemical measurements is how a protein will find its target site and avoid unspecific binding within the millions of base pairs that typically make up a eukaryotic genome (*Stanford et al., 2000*). In spite of these drawbacks, many of the binding parameters obtained with DNA-binding proteins that have been isolated and characterized biochemically can be confirmed by in-vivo measurements.

4.2.1. Structural Aspects of Protein DNA-Interaction

Once a DNA-binding protein has been isolated and purified, it can be characterized by structural methods such as NMR or x-ray crystallography. Such measurements have revealed that there are numerous ways a protein can recognize and bind to a specific DNA-sequence (*Jones et al., 1999; Luscombe et al., 2000*). In accordance with the polar nature of the surface of the DNA molecule, hydrogen bonds, non-polar interactions and buried water molecules are more important in protein-DNA interaction than in protein-protein interactions. Structurally the DNA-binding surfaces of different proteins have little in common; they can consist of α-helices, β-sheets and loops and contact the major and the minor grooves in various combinations. The binding modes can vary between contiguous amino acid residues contacting the backbone atoms and bases of a contiguous stretch of DNA (single headed), two separate clusters of amino acid residues contacting two separate stretches of DNA (double headed) and the enveloping binding mode. This last binding mode occurs predominantly in proteins that perform enzymatic modifications of the DNA, the DNA is bound in a deep cleft of the protein resembling the catalytic pocket of enzymes that use small molecules as substrates (*Jones et al., 1999; Luscombe et al., 2000*). Although DNA that is bound to protein often diverges from canonical B-DNA conformation, the double helix usually remains intact. This is possible because the major and the minor groove of intact DNA present a pattern of hydrogen bond acceptors and donors that is highly typical of its sequence (*Stryer, 1990*). The amino acid residues may contact the DNA through the major groove only, the minor and the major groove, or only the minor groove. The latter has only been observed in complexes in which the DNA is severely distorted from its canonical B-structure (Jones, van Heyningen et al. 1999). It has been proposed that some of the proteins that bind to distorted DNA merely stabilize a DNA conformation that is already energetically favored due to the specific structural properties of the bases that constitute the recognition sequence (*Hagerman, 1990; Travers, 1992*).

In spite of the great diversity of DNA-binding domains, six structural/functional groups of DNA-binding proteins that have a prominent structural feature for DNA-recognition in common can be identified; DNA-binding enzymes constitute their own group. These groups of proteins can be subdivided into homologous families (*Luscombe et al., 2000*).

Helix-Turn-Helix Motif (HTH): These motifs consist of a recognition helix (helix 2) that binds in the major groove and another (helix 1) that is connected with helix 2 by a loop forming a structure of two perpendicular helices. Helix 1 helps to position helix 2 and like the loop, it makes supporting non-specific contacts with the DNA backbone. This recognition motif occurs in prokaryotic proteins that bind the DNA as homodimers; eukaryotic proteins using the HTH motif usually bind the DNA as monomers or as heterodimers contacting asymmetric target sites (*Luscombe et al., 2000*).

Zinc-Coordinating Proteins (ZCP): In eukaryotes, most of the transcription factors are zinc coordinating proteins. They feature one or two zinc ions that are coordinated by conserved histidine and cysteine residues, which are arranged in tetrahedral geometry. The coordinated metal ions help to confer structural stability to domains that are too small to form a conventional hydrophobic core. Like the HTH domain, the ZCPs bind and recognize the DNA by inserting an α-helix into the major groove. Except the ββα-zinc finger proteins all ZCPs form homodimers, heterodimers or tetramers. The ββα-zinc finger proteins have a modular structure, meaning that they are arranged as a covalent tandem of two, three or more separate ββα-zinc finger motifs (*Luscombe et al., 2000*). Because of this simple and modular architecture zinc-finger proteins have been used as the framework for designed DNA binding proteins (*Uil et al., 2003*).

Zipper-Type Proteins: The zipper type proteins are named after their dimerization domain, the leucine zipper. They can be subdivided into two families, the leucine zipper family (LZ) and the helix-loop-helix (HLH) family. In the bound conformation, the members of the LZ family appear as dimers of long helices. The amino-terminal part of each helix is called the basic region and contacts the DNA. Interestingly this region is disordered in solution. Similarly, some HLH proteins have been shown to have a DNA-binding region that is disordered when not bound to DNA (*Phillips, 1994*). However, the DNA-binding region of HLH is separated from the dimerization region by a loop. This allows more flexibility in the positioning of the DNA-recognition helices (*Luscombe et al., 2000*). The zipper proteins show very specific dimerization characteristics with homo- and heterodimers being formed in

various combinations and differing stabilities within the groups. The combinatorial properties of this system allows information about the external and internal conditions to be integrated resulting in specifically adapted patterns of protein expression.
Seven more families employ α-helices as DNA recognition elements; they are functionally and structurally diverse and cannot be grouped.

β-Sheet Proteins: This group consists only of the TATA-box binding proteins (TBP) family. Their role in transcription initiation is best understood for transcription by RNA polymerase II. Unlike the other domains described so far the TATA box binding proteins contact the minor groove. A ten-stranded β-sheet sits on the minor groove in a structure and resembles a molecular saddle. The residues of the TBPs contact the DNA backbone and the minor groove, which is widened by the insertion of four phenylalanine residues (Burley 1996).

β-Hairpin/Ribbon Proteins: This diverse group of proteins uses smaller β-sheets or even β-hairpins to interact with the major or the minor groove of the DNA (*Luscombe et al., 2000 105*). When the protein-DNA interaction occurs in the minor groove the intercalation of amino acid residues causes a significant distortion of the DNA as seen with TBP proteins (*Robinson et al., 1998*).

Interstrand/Interloop DNA binding domains: Two families of non-enzymatic DNA binding proteins have been identified that do not use any combination of the common secondary structure elements but the loops connecting them. These proteins form multimeric complexes that envelop the DNA (*Luscombe et al., 2000*).

DNA-Modifying enzymes: These proteins represent a functional rather than a structural group and comprise enzymes that cleave the DNA at specific sites (endonucleases), enzymes that introduce small chemical modifications of the DNA (i.e. methytransferases) and enzymes such as polymerases that bind to the DNA in a sequence independent manner but still recognize mismatched base pairs. These proteins usually bind DNA in a U-shaped cavity that resembles a substrate-binding pocket and features an active site. In spite of these similarities, the DNA binding domain of DNA-modifying enzymes cannot be described in common structural terms (*Luscombe et al., 2000*).

4.2.2. Thermodynamic Aspects of Protein-DNA Interactions

Techniques such as NMR and X-ray crystallography can be used to examine the protein, the DNA and complexes between protein and DNA. However, the images obtained by these methods do not provide quantitative information about the binding reaction. The single most important parameter for quantifying binding in biological systems is the Gibb's free energy of the binding reaction, ΔG. It expresses the difference between two energetic states (G) of a system. In the context of the binding reaction, this is the difference in energy between free DNA and free protein and the protein-DNA complex. Therefore ΔG describes how energetically favorable the binding, how high the affinity between protein and DNA, is. This affinity is measured in kJ mol^{-1}, when it is negative the binding energy is favorable and the energy of the complex is lower than of the separate states. Thus, the considered molecules will form a complex when mixed under the considered conditions. Conversely separating the complex will require an energetic input of ΔG kJ mol^{-1}.

In order to understand ΔG, the difference in free energy of two systems, it is necessary to understand G, the free energy of a given system. Gibb's free energy (G) is a composite number that is defined by:

$$G = H - TS \tag{1}$$

H is the enthalpy of the system, again a composite number with the unit kJ mol^{-1} that is the sum of the internal energy and the product of the pressure of the surroundings and the volume of the system:

$$H = U + pV \tag{2}$$

T is the absolute temperature in Kelvin and S is the entropy of the considered system. Entropy is usually interpreted to be the probability of the system, high entropy indicating high probability and low entropy indicating low probability. Note that order is often used synonymously with probability. Entropy is usually given as relative entropy, ΔS, the logarithmic function of the ratio of the probability of the respective states (*Alberts, 2002*):

$$\Delta S = R \ln \frac{probabiliy\ \left(state1\right)}{probabiliy\ \left(state2\right)} \tag{3}$$

In order for the binding reaction to occur, ΔG must be negative. Because ΔG is a composite number defined by:

$$\Delta G = \Delta H - T\Delta S \tag{4}$$

binding is favored when the complex has a either a lower enthalpy or a higher entropy than the free DNA and the free protein. As long as ΔG is negative, the binding reaction occurs. In this way a favorable decrease in enthalpy can drive a reaction with an unfavorable decrease in entropy and vice-versa. Sources of favorable, meaning negative, ΔH contributions in protein-DNA binding are the recognition contacts, hydrogen bonds, hydrophobic interactions and salt bridges between protein and DNA. There are several potential sources of unfavorable, that is positive ΔH contributions. Firstly, there is the partial or complete desolvation of polar surfaces as the interaction interfaces are brought together. Secondly, the binding molecules may experience conformational strain in the complex. This occurs when in the complex bonds have to adopt conformations that do not coincide with the conformations of minimal energy these bonds would adopt in an uncomplexed molecule. Naturally, this kind of strain occurs only when the net loss of free energy favors it. An obvious source of strain is the distortion of DNA that often occurs when the DNA is bound by protein (*Jen-Jacobson et al., 2000*). The intermolecular interactions, dynamic restrictions and the reorganization of buffer components that contribute to ΔH also determine ΔS. One favorable, meaning positive, contribution to ΔS is the release of water molecules from the binding interface, especially from the non-polar surfaces. Unfavorable contributions to ΔS are the loss of translational and rotational freedoms of protein and DNA (determined by the number of elements joined) and the reduction of vibrational and conformational entropy (*Jen-Jacobson et al., 2000*). Experimentally only ΔG and ΔH are accessible and ΔS must be deduced from these values and the temperature in Kelvin. There is no direct experimental way to determine the different contributions to ΔH and ΔS. Looking at a number of protein DNA complexes it appears that distortions of the DNA are associated with entropically driven reactions with unfavorable enthalpy, and complexes where the DNA appears to be in its regular B-conformation are associated with enthalpy-driven reaction with unfavorable entropy (*Jen-Jacobson et al., 2000*).

4.2.3. Kinetic Aspects of Protein DNA-Interactions

The measurement of the equilibrium concentrations of DNA and protein and DNA-protein complex can define the relative thermodynamic properties of the free and the bound state of the molecules, but information about the pathway between these states can be derived only by non-equilibrium measurements (*Fersht, 1998*). Thermodynamics describe the states of a system rather than how states come into being. However, for all practical purposes how and within which timeframe a system comes into being is a very crucial question.

It has been recognized that the cell nucleus and its components are a highly dynamical system enabling a prompt response to signaling cascades caused by the external or internal conditions a cell encounters (*Misteli, 2001*). In order for sequence specific protein-DNA interaction to help mediate the response to a change in external or internal conditions the interaction must occur promptly. How can a specific recognition process occur promptly when the recognition protein will only rarely collide directly and precisely with the cognate site (*Halford and Szczelkun, 2002*)? This is only possible if the protein can move from one location to another within the DNA polymer on a biologically useful timescale. Therefore, kinetic measurements are necessary to understand how a specific protein-DNA complex develops. A kinetic measurement is itself a real-time observation of how a system reaches equilibrium. In studying protein-DNA interactions, kinetic measurements are used to determine the rate of formation and the rate of dissociation of the complex. This is usually accomplished by tracing the increase in product concentration or the decrease in reactant concentration. Often products and reactants have markedly different spectroscopic properties. Time resolved changes in spectroscopic properties obtained by appropriate stopped flow methods can be correlated to changes in product and reactant concentration.

One problem that proteins have to overcome to achieve sequence-specific DNA binding is locating the cognate binding site among the millions of base pairs that typically make up a genome. Unspecific binding may on one hand facilitate finding the target sequence by allowing one-dimensional diffusion along the length of the DNA, on the other hand the formation of overly stable complexes with non-cognate sequences, i.e. getting “stuck”, must be avoided (*Stanford et al., 2000*). Unspecific binding is arbitrarily defined as binding to random sequences, which occurs with much lower affinity than to the cognate site. Unspecific binding can be further discriminated from specific binding because it does not occur with the stable stoichiometry predicted by the number of recognition sites. Thermodynamically, the less favorable ΔG of non-specific binding originates from a smaller negative ΔH. Some measurements using ITC (Isothermal Titration Calorimetry) even indicate that non-specific binding is driven by a favorable positive entropy change (*Takeda et al., 1992*). The importance of long-range electrostatic interactions, rather hydrogen bonds and van der Waals contacts, in unspecific binding can be inferred from its higher sensitivity to the salt concentration in the buffer. The ΔCp upon non-specific binding is less negative than upon specific binding because less water is released from the binding surfaces that are juxtaposed less tightly. Interestingly, the kinetic basis of the smaller affinity of DNA binding proteins to

random sequences is a higher dissociation rate constant (*Oda and Nakamura, 2000*). The higher dissociation rates from non-specific sequences indicate that the transition state for binding/dissociation is more accessible from a non-specific complex than from a specific complex. These observations from biophysical measurements accord very well with the biological role of unspecific binding. The first encounter between the DNA polymer and the protein generally is unspecific binding, which is dominated by long-range interactions that direct the protein into the sphere of the DNA leading to a high association rate. If an unspecific complex results from this encounter, it dissociates rapidly and can re-associate at another topologically close position on the same DNA polymer.

Although the role of one-dimensional diffusion has been confirmed by numerous experiments, involving the binding rate of protein to DNA targets of various lengths, its role in vivo remains questionable because of the obstacles that complicate sliding motions on the DNA over any significant distance (*Gowers and Halford, 2003*). The DNA is usually packed as euchromatin, which consists of DNA bound to proteins called histones, which form a repeating array of protein-DNA particles called nucleosomes. Stretches of euchromatin are interrupted by heterochromatin, an even more packed form of DNA that interacts with additional proteins and often is resistant to expression (*Alberts, 2002*). Three-dimensional diffusion, where DNA dissociates rapidly from non-cognate DNA and rebinds to another site on the same macromolecule is much less affected by the circumstances described above. An elegant experimental approach for comparing the roles of one-dimensional and three-dimensional diffusion involves measuring the competitive cleavage of plasmid, catenane and mini-circle DNA by a restriction endonuclease, while keeping DNA concentration and ratio of cognate cutting site to non-specific DNA constant (*Gowers and Halford, 2003*).

4.3. Experimental Characterization of Protein-DNA Interaction

4.3.1. Measuring Thermodynamic Parameters for Protein-DNA Interaction

The most accessible thermodynamic parameter of protein DNA binding is ΔG. As elaborated previously ΔG quantifies the affinity of the protein to its target sequence. A perhaps more descriptive way to quantify the affinity of a protein to its target sequence is the equilibrium constant.

Considering the following reversible reaction where P is protein, D is DNA and P·D is a specific complex between protein and DNA:

$$P + D \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PD$$

the association constant K_A is defined as:

$$K_A = \frac{[PD]}{[P][D]} \tag{5}$$

while the dissociation, constant K_D is the reciprocal value of K_A. The free energy of binding is related to the association constant by:

$$\Delta G = -RT \ln \frac{[PD]}{[P][D]} \tag{6}$$

ΔG equals zero when:

$$[P][D] = [PD] \tag{7}$$

This means that the reaction is at equilibrium and there is no driving force in either direction. In this state:

$$k_{on}[P][D] = k_{off}[PD] \tag{8}$$

Clearly, ΔG is a function of the concentration of the reactants and products. Usually the affinity of two molecules is indicated by the standard free energy change ΔG^0. The standard free energy change depends on the intrinsic properties of the reacting molecules and is the free-energy difference under standard conditions. Ambient standard conditions are defined as 298.15 K, 101.3 kPa and all components of the reaction present at 1mole l^{-1} (1M). The ΔG with other concentrations of reactants can be calculated according to:

$$\Delta G = \Delta G^0 + RT \ln \frac{[PD]}{[P][D]} \tag{9}$$

Mixing the reactants in different ratios and determining the concentration of the complex can measure the free energy of a binding reaction. Normally the concentration of one of the reactants A is kept constant while the concentration of the other reactant B is increased by adding small aliquots of a very concentrated solution of B. The concentration of the complex can be monitored spectroscopically if it has different fluorescence or circular dichroism properties from either of the reactants. In the case of protein binding to DNA, only proteins and only those containing tryptophan or tyrosine residues show an intrinsic detectable

fluorescence. The fluorescence yield of these amino acids is highly sensitive to the polarity of the surrounding solvent. If the binding reaction causes a change in the solvent environment of the fluorophores, this will cause a change in fluorescence yield. If the maximum change in fluorescence yield, i.e. the change at saturation is known, the observed change can be used to quantify the fraction of bound and free protein at any ratio of reactants. More often, the intrinsic fluorescence of the protein is quenched upon binding to DNA. In addition, here the fraction of bound and free protein at any point of the titration can be determined if the extent to which the protein fluorescence is quenched when fully bound to DNA is known. The fluorescence of the DNA is generally negligible. In order to obtain a reliable binding constant the concentration of the reactant that is being held constant should be smaller than K_D. If this is not the case, there will be little free reactant at stoichiometric concentration of protein and DNA. In the case of proteins with very high affinity to their target DNA this behavior may lead to significant problems regarding the fitting to a binding isotherm (*Moss, 2001*).

Not all proteins that bind to DNA have intrinsic fluorescence, and not all proteins that have intrinsic fluorescence show a fluorescence change upon binding to DNA. In such cases, it can be useful to modify the DNA with a fluorescent reporter molecule that changes its quantum yield or its emission/excitation maximum when the DNA is bound by protein.

4.3.2. Measuring Kinetic Parameters for Protein-DNA Interaction

The biological implications of a reaction do not only depend on its equilibrium properties. The timescale on which the reaction occurs is of equal importance. Note that all reactions taking place in a cell are net thermodynamically favorable and should occur spontaneously. Enzymes catalyze reactions that would take place immeasurably slowly on their own; frequently they also couple energetically unfavorable reactions to reactions that are favorable. As described previously a kinetic measurement follows the increase of a product or the decrease of concentration of a reactant as a function of time after the two reactants have been mixed. If the reaction proceeds through a single rate liming state, the decrease of reactant or the increase of product can be described by a single exponential equation. Reactions proceeding over several rate-limiting barriers are described by a sum of exponentials. For any binding reaction at least two parameters can be measured, the binding rate k_{on} and the dissociation rate k_{off}:

$$P + D \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PD$$

For a second-order binding reaction, which depends on the concentration of both reactants, k_{on} is measured in $M^{-1}s^{-1}$ and k_{off} is measured in s^{-1}. The analysis of this system is difficult because the observed increase in product depends on the concentration of both reactants.

$$\frac{d[PD]}{dt} = k_{on}\left([P_0]-[P]\right)\left([D_0]-[D]\right) - k_{off}[PD] \tag{10}$$

$[P_0]$ and $[D_0]$ are the initial molar concentrations of protein and DNA, $[P_0]$-$[P]$ and $[D_0]$-$[D]$ are the concentrations of protein and DNA at each time point. The concentration of the product at each time point is therefore the integral of equation (9):

$$\left[PD(t)\right] = [P_0]-\left[P(t)\right] = [D_0]-\left[D(t)\right] \tag{11}$$

The advantage of this type of data analysis is that the dissociation and the association rate constant can be obtained from a single binding trace. The drawback is that binding traces of high quality need to be recorded for the parameters to be sufficiently constrained.

Another method for obtaining the rates of association and dissociation can be used when the equilibrium stability of the complex is known. The equilibrium concentration of each species at infinite time can be calculated and a linear equation can be used to describe the rate constants of the reaction as a function of the concentrations of protein and DNA at equilibrium (*Bernasconi, 1976*):

$$k^{app} = k_{on}\left(\left[\overline{P}\right]+\left[\overline{D}\right]\right)+k_{off} \tag{12}$$

The terms $\left[\overline{P}\right]$ and $\left[\overline{D}\right]$ stand for the concentration of the respective species at equilibrium.

The apparent rates for different equilibrium concentrations of reactant are measured and plotted against the sum of the equilibrium concentration of protein and DNA.

When the kinetic trace of a binding reaction only fits to the sum of two exponentials rather than to a single exponential a more complicated model is needed to describe the binding. One possibility is association through an intermediate:

$$P + D \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} PD^* \underset{k_{off}'}{\overset{k_{on}'}{\rightleftarrows}} PD$$

The intermediate species PD^* converts to the final product in a first order transition. If the amplitudes of the two observed rates k_{app} and $k_{app'}$ are large enough all four binding constants that describe this system can be calculated. Equation 11 is used to determine the rate of

association and dissociation and equation 12 to determine the rate of the rearrangement that the complex undergoes after association or before the dissociation, respectively (Bernasconi 1976).

$$k^{app}{}' = k_{on}{}' \frac{k_{on}\left(\left[\overline{P}\right]+\left[\overline{D}\right]\right)}{k_{on}\left(\left[\overline{P}\right]+\left[\overline{D}\right]\right)+k_{off}} + k_{off}{}' \qquad (13)$$

The value of the dissociation rate k_{off} or the rate of the rate-limiting step in dissociation can be determined in isolation from binding by a displacement experiment. This is convenient when the reaction is followed by monitoring the fluorescence change of a labeled DNA upon binding to protein. Pre-equilibrated complex of fluorescently labeled DNA and protein is mixed with a significant excess of unlabeled DNA. Consequently, every protein that dissociates from the labeled DNA has a very high chance of rebinding to unlabeled DNA. The latter does not lead to an observable signal. The disappearance of the characteristic signal of the complex between protein and labeled DNA is thus a direct measure of the rate-limiting step of dissociation.

$$PD_{labelled} \underset{k_{on}}{\overset{k_{off}}{\rightleftharpoons}} P + D_{labelled} \xrightarrow[k_{on} >> k_{off}]{\text{+excess of unlabelled DNA}} PD_{unlabelled}$$

The length of the arrows in the reaction scheme are arbitrary and merely emphasize that a protein molecule that has once dissociated from a labeled DNA has a much higher chance of rebinding to unlabeled DNA than of rebinding to labeled DNA.

4.4. bZIP Transcription Factors

As described previously the DNA recognition domain of zipper-type proteins consist of a dimerization region called leucine zipper and a DNA-binding sequence called basic region. The DNA recognition domain in leucine zipper proteins consist of two extended helices, the carboxy terminal part of each of the helices functions as a dimerization domain, the amino terminal part, also called basic region, interacts with the DNA (Figure 4.1). The architecture of the DNA recognition domain of helix loop helix proteins is similar but a loop is inserted between the dimerization and the DNA-binding domain. Zipper-type proteins are found widely in eukaryotic organisms as diverse as mammals, birds, insects, higher plants, fungi, and slime molds. On the most basic level the function of bZIP transcription factors is to bind to DNA in a sequence specific manner and enhance the efficiency with which RNA polymerase II binds to and initiates the transcription of a certain gene. Some bZIP

transcription factors enhance translation constitutively while the activity of other bZIP transcription factors depends on post-translational modification that occurs in response to external stimuli (*Hurst, 1996*).

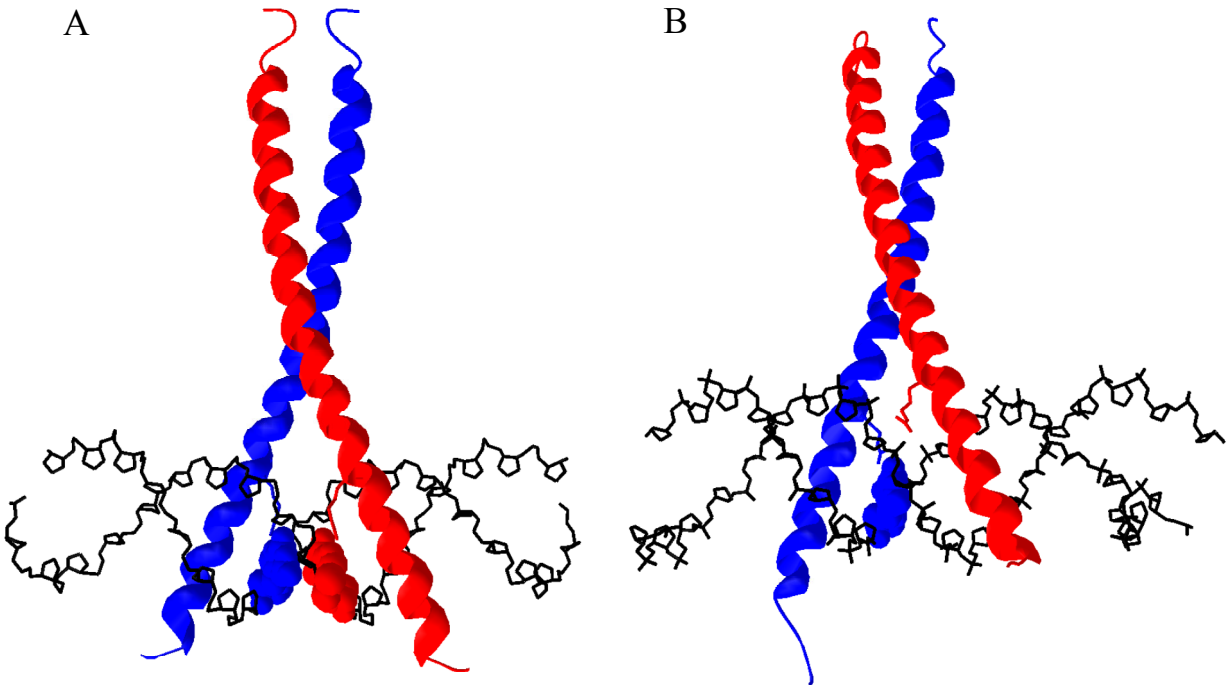


Figure 4.1 A fragment of the transcription factor GCN4 which is sufficient for DNA binding and dimerization crystallized in complex with, A: the palindromic recognition site ATF/CREB where the arginine 243 of each chain forms a contact to the central guanines (*Konig and Richmond, 1993*), B: the pseudopalindromic AP-1 sequence where only the arginine of the left chain forms a base recognition contact with the single central guanine (*Ellenberger et al., 1992*). The two protein chains are depicted as a red and a blue ribbon. Arginine 243 which forms a contact either to the backbone or to the central guanine is shown as a ball and stick moiety. The backbone of the DNA is shown in black; the central guanine(s) are shown in color and as a space-filled moiety.

4.4.1. Dimerization

Although the dimerization domain is not directly involved in DNA binding it plays a large role for the biological function of bZIP transcription factors. The only conserved feature of the dimerization domain is the periodic occurrence of leucine every seventh amino acid (position 1, 8, 15, 22), or every second helical turn. This arrangement results in a strip of leucine residues that wind around the outside of the helix as seen in Figure 4.2. Closer examination of the amino acid sequence reveals that hydrophobic residues are often found at positions n+4/ n-3 (5 and not shown 12, 19, 26) relative to the conserved leucine.

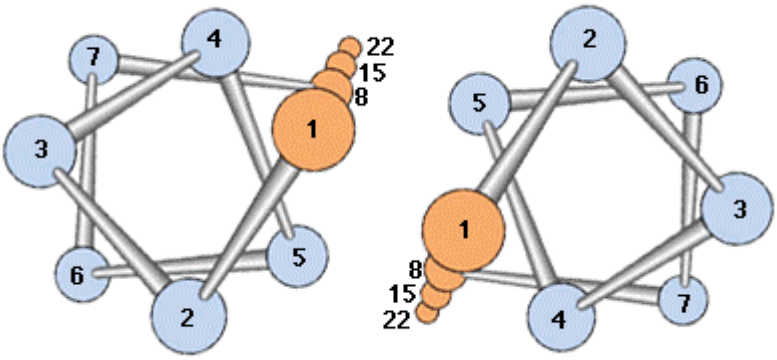


Figure 4.2 Schematic helical wheel depiction of a coiled-coil domain. In a leucine zipper the positions shaded in pink are mainly occupied by leucine. The position labeled 5 is usually occupied by a hydrophobic residue. The positions labeled 2 and 4, adjacent to the dimerization interface are often occupied by polar amino acid residues.

The strip of leucine residues found on the outside is therefore part of a two-residue wide strip of hydrophobicity. Only the hydrophobic n+4/ n-3 positions toward the center of the leucine zipper are sometimes occupied by a polar residue. This hydrophobic strip is the dimerization interface of the bZIP transcription factor; as a result, the two leucine zippers have to wind around each other to dimerize, thus the supercoiling observed in leucine-zipper type coiled coils. Positions adjacent to the hydrophobic strip n+1/n-7 (2, and not shown 9, 16, 23) and n+3/n-4 (4, and not shown 11, 18, 25) are usually occupied by charged amino acids. Together with the polar residues that are sometimes found in the center of the dimerization interface, these residues determine the interaction specificity of the leucine zipper (*Hurst, 1996*). Selective formation of homo- and heterodimeric protein complexes generates a high level of DNA binding diversity based on relatively few proteins (*Wolberger, 1999*). Although all PDB entries for bZIP proteins show homodimers, heterodimeric pairing has been observed biochemically and its biological effect analyzed (*Vinson et al., 2002*). Recently a protein array technique has been developed which tests the interaction of 49 human bZIP transcription factors, which could give rise to 492 pairings. This technique identified a number of new pairings between domains that occur in the same tissues, at the same time and that have been linked to the same processes and probably have biological importance. These findings highlight that there is still much to be learnt about the role of bZIP dimerization (*Newman and Keating, 2003*).

4.4.2. DNA-Binding

At the amino terminus of the leucine zipper, the two interacting α-helices diverge and enter the DNA major groove in opposing directions. The part of the α-helical DNA-binding domain

of bZIP transcription factors that contacts the DNA is called the basic region. Usually the basic region is located C-terminally of the leucine zipper although it can also be located toward the N-terminus of the protein (*Hurst, 1996*). Only few positions are conserved and mutational analyses and examination of the structure indicates that these conserved positions are important for binding (*Hurst, 1996*). When the first residue after the last leucine is designated -1 and the basic region residues numbered accordingly, the most conserved positions are an arginine at position -10, an alanine at position -14 or -15 and an asparagine at position -18 (Figure 4.3) (*Hurst, 1996*).

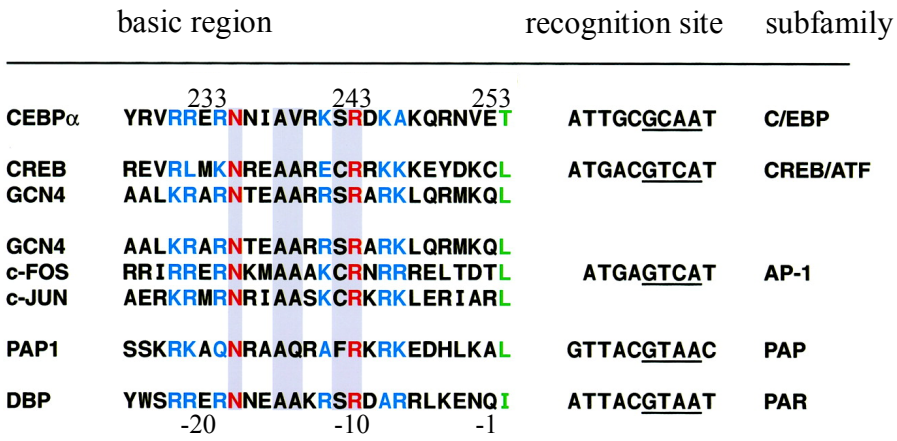


Figure 4.3 Sequence alignment of the basic regions of representative members of five subfamilies of the bZIP proteins. The strictly conserved residues are depicted in red, moderately conserved basic residues in blue and the first leucine of the leucine zipper in green. Other residues shaded in grey are also moderately conserved. The numbering above refers to the complete GCN4 protein sequence. The numbering on bottom refers to the first residue after the last leucine as -1 and counts backwards using negative numbers to the amino-terminus of the basic region. Figure according to (*Miller et al., 2003*), modified.

The basic regions of homodimeric bZIP transcription factors usually recognize palindromic sites, where each arm of the dimer contacts one half-site resulting in a symmetric structure. Some bZIP transcription factors recognize a pseudopalindromic site, which means that the two recognition sites overlap at the central base pair. Many bZIP proteins bind to various palindromic sequences with differing affinity. Therefore bZIP proteins are promiscuous in their DNA-binding as well as in their dimerization behavior and their affinity and their specificity is modulated not only by whether the proteins is dimerized or not but also by which two proteins are taking part in the dimer. Apart from these factors, derivatization

(usually phosphorylation) and interaction of a dimer with other proteins can affect binding affinity and specificity (*Hurst, 1996*).

Examination of the protein-DNA complexes of GCN4, Fos and Jun and the pseudopalindromic AP-1 sequence reveals that residues that are conserved in the entire bZIP family contact specific bases, implying that the DNA-binding sites of bZIP proteins should have these specific bases in common. However, the recognition sites for the members of the bZIP transcription factor family differ and some do not have the bases that make specific contacts to the conserved amino acid residues (*Fujii et al., 2000*). The conserved residues and their contacts to bases, as described below, probably help to provide a scaffold for other interactions rather than conferring specificity themselves. The central base pair/s of the palindromic or pseudopalindromic recognition sites are usually G/C pairs. The guanine of the base pair is contacted by the conserved arginine 243 (-10) by way of hydrogen bonds that are formed between the two guanidine-groups of the arginine as hydrogen bond donors and the 7-N and 6-O groups of the guanine as acceptor. In the ATF/CREB site, where two guanines occur, the contacts between the protein and the bases are totally symmetrical. When only one central base pair occurs as in the AP-1 site only the left monomer establishes a contact to the central guanine while the other makes hydrogen bonds to the phosphate backbone. Numbering the DNA-binding site from the central guanine/s outwards, position 2 is occupied by adenine or cytosine, it follows that position 2' is occupied by guanine or thymine. When a thymine occurs, as in ATF/CREB and AP-1, it can be recognized by alanine 238 (-15) through a hydrophobic contact between the methyl-group of the alanine and the 5-C methyl-group of the thymine. At position 2 a cytosine occurs in the ATF/CREB and AP-1 sites, other recognition sites feature an adenine in this position. The adenine/cytosine is contacted by the conserved asparagine 235 (-18), the γ-carboxylic group of which functions as a hydrogen bond acceptor and the 6-N or 4-N residues of the A or C, respectively, acts as hydrogen bond donor. At position 3 the consensus base is adenine with a thymine occurring at position 3'. This thymine forms a contact with the conserved alanine residue at position 239 (-14) of the basic region as well as with serine 242 (-11). It also interacts with asparagine -18 by way of a hydrogen bond between the 4-O of the base and the γ-amide group of the amino acid. Not shown in Figure 3.4 are the contacts which arginines 249 (-4), 245 (-8), 241 (-12), 240 (-13), 234 (-19) and 232 (-21), lysine 246 (-7) and threonine 236 (-17) make to the DNA-backbone. Interestingly the contacts made to the backbone are asymmetric in the AP-1 complex, while direct base contacts (with the exception of the central guanine) are symmetrical and identical to those observed in the ATF/CREB complex (*Ellenberger et al., 1992; Hurst, 1996; Keller et*

al., 1995; Konig and Richmond, 1993). It thus follows that the two half-sites and the two monomers, respectively, may be of different importance in binding to AP-1.

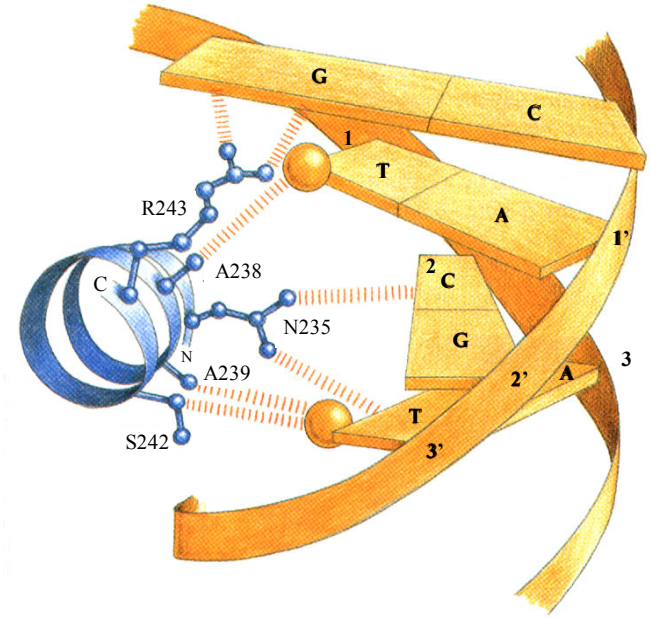


Figure 3.4 Detailed schematic depiction of protein-DNA interactions at the recognition-surface of GCN4 and AP-1. Only direct contacts between the bases and amino acid residues are shown. Contacts of the protein to the backbone have been omitted for clearness. Figure is taken from (*Branden, 1998*).

Sequence alignments between different bZIP transcription factors and mutagenesis experiments indicate that residues that form contacts to the phosphodiester backbone may be more important for specificity than the most conserved residues and their interactions with specific bases as seen in Figure 3.4. Naturally, residues that bind to the backbone must recognize the DNA conformation, which may be the result of combinations of specific bases rather than the specific bases themselves. Similarly, shifts in position and orientation on the DNA and local deformation of the α-helical recognition region, rather than specific base contacts made by certain amino acid residues can modulate binding specificity of the bZIP proteins (*Kim et al., 1993*). Furthermore, the variable residues can change the conformation of the conserved residues, which may then interact with the bases in various ways dependant on sequence context (*Fujii et al., 2000*).

An important feature of DNA binding by bZIP transcription factors is induced distortion of the DNA, especially in the context of global recognition strategies based on compatible

conformations rather than specific amino acid residues recognizing bases. When GCN4 is bound to the palindromic recognition sequence ATF/CREB that is one residue longer than AP-1, the conformation of the DNA diverges from the usually observed B-DNA. The protein-DNA contacts in this complex are very similar to the contacts observed when GCN4 is bound to AP-1. This is possible on one hand because the slightly bigger divergence of the subunits at the fork where the α-helices enter the large groove and on the other hand because the DNA is distorted; in fact the central region of the recognition complex adopts a structure reminiscent of A-DNA (*Keller et al., 1995*). According to solution studies, the ATF/CREB sequence is bent even when not bound by protein (*Khandelwal et al., 2001*). This confirms the idea that the distortion of the DNA observed in the complex is partly the result of further stabilization of a state that is energetically favored to begin with. DNA bending in general plays an important role in DNA-binding by GCN4. Many members of the bZIP family of transcription factors either bind to bent recognition sites or bend the recognition sequences they bind to. Whether the DNA in a complex with a bZIP transcription factor is bent depends not only on the recognition site itself, but also on the combination of transcription factors that are bound to the studied site.

The mechanism by which a bend in the DNA is stabilized or introduced has been studied experimentally by mutagenesis and chemical modification of the protein. The results indicate that when DNA is bent toward protein this is caused firstly by strongly favorable protein-DNA interactions that can be maximized by the bend and secondly by salt bridges if they are formed preferably on one side of the helix. This leaves the interphosphate repulsions on the other side of the helix unbalanced and the DNA relaxes toward the neutralized surface. In support of this it has been observed that many proteins that bend the DNA toward themselves contain cationic residues (Tomky 1998). The direction of the bend depends on the precise location of the cationic residues; this mechanism can thus also explain a bend of the DNA away from the main body of the protein. It remains unclear whether the basic residues bend DNA by forming salt bridges with the phosphate back bone in an orientation which requires the DNA to bend, or whether the bend is the result of asymmetrical neutralization of charges on the DNA. Anionic residues can also cause an asymmetric charge distribution on the DNA; the effects are converse of what is described above. When DNA is bent away from the protein this can be due to intercalation of hydrophobic residues between adjacent base pairs in the minor groove thus enlarging this group and distorting the helix axis. Furthermore, theoretical studies indicate that DNA is bent away from protein because the side of the DNA that is interacting with the protein experiences a medium with a lower dielectric constant. The

electrostatic repulsions on the bound face of the helix are therefore stronger and the helix relaxes away of the area of lower dielectric constant (*Maher, 1998*).

4.5. Investigating the DNA Binding of C62GCN4

One of the best characterized bZIP proteins is the yeast transcription factor GCN4. In *Saccharomyces cerevisiae* GCN4 functions as a positive regulator of transcription that controls the biosynthesis of multiple enzymes that occur in the biosynthetic pathways of amino acids. The expression of GCN4 itself is triggered by the shortage of amino acids and/or other nutrients in the growth medium of the yeast and is regulated on the translational level by GCN2 (*Hinnebusch, 1984*).

New studies employing cDNA microarray analysis reveal that more than 1000 genes are induced in response to histidine starvation and an equal number is repressed by a factor of two or more. Using a GCN4 deletion mutant it could be shown that at least 539 genes are targeted by GCN4 only 25% of which are involved in biosynthetic pathways of amino acids. Further analysis of the involved genes shows that numerous protein kinases and transcription factors are under GCN4 transcriptional control suggesting that GCN4 is a master regulator of gene transcription (*Natarajan et al., 2001*). The same study could also show that GCN4 can be induced by treatment with the alkylating agent methylmethanesulfonate indicating that GCN4 expression is also part of a general stress response. Experiments characterizing the biophysical properties of GCN4 are often carried out using a fragment comprising approximately 60 C-terminal residues of GCN4. Such fragments are sufficient for DNA-binding and dimerization as they encompass the leucine zipper as a dimerization domain and the basic region, which is responsible for DNA-binding. In these constructs the transactivation domain that is necessary for GCN4 to fulfill its biological function is missing (*Hurst, 1996*).

4.5.1. Dimerization of C62GCN4

GCN4 is usually observed as a dimer binding to a recognition sequence of dyad symmetry. Although GCN4 can also bind protein as a monomer, the affinity of a monomer to DNA or of a dimer to a recognition half site is much smaller than the affinity of the dimeric transcription factor to the full recognition site (*Hollenbeck and Oakley, 2000; Wang et al., 2003*). Thus it is necessary to understand the dimerization properties of GCN4 to understand its DNA-binding properties. Experiments to elucidate the dimerization properties of GCN4 can be done using a

fragment that corresponds to the leucine zipper element only, because the basic region and the transactivation domain are not necessary for dimerization. Using a denaturant the coupled folding and dimerization, or coupled unfolding and dissociation of such a fragment was observed by stopped-flow circular dichroism spectroscopy. Such experiments show that denaturant induces a cooperative transition between folded dimer and unfolded monomer which occurs with $\Delta G = 10.5$ kJ mol^{-1} dimer at 5 °C and pH 7.0. The rate of association and folding is 4.18×10^5 $M^{-1}s^{-1}$ and the rate of dissociation and unfolding is 3.34×10^{-3} s^{-1}. The rather small rate of association may indicate that productive collisions occur only between unstable monomeric intermediates that are partially helical (*Zitzewitz et al., 1995*). The formation of the coiled coil of GCN4 has also been studied by following the fluorescence increase of a tryptophan that was introduced in place of a tyrosine found in the natural GCN4 leucine zipper. Upon dimerization this tryptophan becomes buried and its fluorescence increases by 50%. Again, the association and dissociation reaction of the monomers and the dimer, respectively, were prompted by dilution from high to low or low to high urea concentration, respectively. The kinetic folding reaction of several mutants was investigated in order to elucidate if helix formation has to occur prior to association or can occur after a productive collision and subsequent collapse (*Sosnick et al., 1996*). However, the thus obtained results allow controversial interpretations concerning the importance of nascent helicity in the monomers for dimerization (*Myers and Oas, 1999*). The importance of preformed helix in the monomers for the dimerization was confirmed by a study that introduced extensive mutations to change the helical propensity over a large range (*Myers and Oas, 1999*). The question of the role of nascent secondary structure in the dimerization of GCN4 leucine zipper was also addressed by studying the effect of SDS on the dimer. It could be shown that while SDS at concentration of 1mM completely dissociates GCN4 leucine zipper dimers, the resulting monomers have a significant α-helical residual structure (*Meng et al., 2001*). However, it remains unclear whether the observed species is really equivalent to the association competent monomer under native conditions.

More detailed analysis of the dimerization of GCN4 leucine zippers are possible using differential scanning calorimetry, chemical denaturation at different temperatures and kinetic measurements of association and dissociation at different temperatures (*Bosshard et al., 2001; Ibarra-Molero et al., 2001*). Employing these methods the activation parameters for folding and unfolding $\Delta G^{\ddagger}$, $\Delta S^{\ddagger}$, $\Delta H^{\ddagger}$ and $\Delta C_p^{\ddagger}$ can be determined. A problematic aspect of characterizing the transitions state thermodynamically is the choice of the pre-exponential factor k_0 in the Eyring equation.

$$k_{X-TS} = \kappa \times k_0 \times e^{\frac{-\Delta G^{TS}}{RT}}$$

This factor reflects the maximal rate of the binding reaction which, assuming that every encounter of two leucine zippers is productive, is limited by the diffusion speed of the reacting molecules. If the pre-exponential factor is for example overestimated, the energy barrier will be overestimated. Because ΔS is calculated from ΔH which can be deduced from the temperature dependence of the rates and ΔG which strongly depends on the pre-exponential factor, the overestimation of the latter will lead to an overestimation of the loss of entropy (the unfavorable contribution ΔS) for reaching the transition state (*Bosshard et al., 2001*). Bosshard et al. and Ibarra-Molero et al. use very similar pre-exponential factors of 10^9 s^{-1} and 3.3×10^9 s^{-1}, respectively, in their studies of GCN4 dimerization. Thus very similar thermodynamic parameters for the self-association of the C62GCN4 construct used by Bosshard et al, which contains the basic region and the isolated GCN4 leucine zipper used by Ibarra-Molero et al. were determined (*Bosshard et al., 2001; Ibarra-Molero et al., 2001*). Regarding the thermodynamic parameters of the transition state the results of these two authors differ somewhat. In the results of Bosshard et al. the strong temperature dependence of the unfolding rate and the small temperature dependence of the folding rate indicate that the main loss of heat capacity, and therefore the main desolvation step occurs only after rate limiting transition state. The experiments furthermore show that the formation of the transition state from two monomers occurs with a small entropic gain which is compensated by a larger loss of enthalpy resulting in an activation barrier of ΔG=12.8 kJ mol^{-1} (at 20° C) (*Bosshard et al., 2001*). Ibarra-Molero et al. also measure the temperature dependence of the enthalpy and energy of the folding and unfolding reaction of the GCN4 leucine zipper. Although they find an energetic barrier of similar magnitude between the unfolded monomeric state and the folded dimer as Bosshard et al, their measurements indicate that formation of the transition state from the monomers is entropically unfavorable by 23 kJ mol^{-1} and enthalpically favorable by 6.7 kJ mol^{-1} and thus occurs with 16.3 kJ mol^{-1}.

4.5.2. Thermodynamic Characterization of DNA Binding of C62GCN4

Thermodynamic data for binding of GCN4 to its recognition sequence is difficult to analyze because many different processes occur at once. Firstly, the binding of GCN4 to DNA is coupled to a folding transition in which the basic region, which is disordered in solution, adopts an α-helical conformation (*Berger et al., 1996; Weiss et al., 1990*). Secondly, GCN4 binds not only to the pseudopalindromic AP-1 site but also to the fully palindromic ATF/CREB site. The binding of GCN4 to an ATF/CREB target site results in a complex in

which the DNA is bent toward the DNA by 20° (*Keller et al., 1995*). It is not known if a kink of this magnitude is inherent to the ATF/CREB sequence or becomes stabilized upon binding of GCN4. Thirdly, the binding of a dimeric transcription factor to DNA is usually considered which may be problematic because monomers can bind the recognition site and conversely a dimer can bind to a DNA containing just half of the consensus half-sites (*Hollenbeck and Oakley, 2000; Wang et al., 2003*). In spite of these problems two experimental studies explicitly characterizing thermodynamic parameters of the interaction between DNA and GCN4 have been published.

One of these studies treats the binding of dimeric GCN4 to the ATF/CREB and AP-1 recognition site by isothermal titration calorimetry. This method allows for the simultaneous detection of the stoichiometry, the association-constant and the binding enthalpy, from which ΔG and ΔS can be derived. To rule out monomer binding the experiments were conducted with protein concentrations at which the monomer concentration of the used constructs is negligible. The results that Berger et al. obtained indicate that C62GCN4 binds to the AP-1 site and the ATF/CREB site with the same affinity (*Berger et al., 1996*). Although both reactions are highly exothermic and show an unfavorable entropy change, the ΔH for binding to the AP-1 site is more favorable and the entropic penalty higher. From the temperature dependence of the enthalpy the change in heat capacity upon binding is calculated. More heat capacity is lost upon binding to the ATF/CREB site than upon binding to the AP-1 site. There is an empirical relation between the loss of solvent accessible polar and non-polar surface area and the change in heat capacity upon binding. In the case of GCN4 binding to ATF/CREB, the loss of heat capacity is larger than predicted by the empirical correlation. Other possible causes of loss of heat capacity are loss of internal flexibility of the DNA and changes in the vibrational term (*Berger et al., 1996*). This study treats binding of the C62GCN4 dimer to the different bipartite recognition sites as a whole. Results obtained by Hollenbeck et al. using quantitative electrophoretic mobility shifts suggest that the two dimers and the two recognition half sites may play rather different roles with one monomer making specific contacts and the other providing binding energy by contacts that are independent of DNA sequence (*Hollenbeck and Oakley, 2000*).

The other study thus focuses on thermodynamically characterizing the binding of monomeric GCN4, to either the pseudopalindromic AP-1 or the fully palindromic ATF/CREB site. Formation of complexes consisting of double stranded DNA with bipartite binding sites and only a single specifically bound DNA molecule was achieved by using a peptide

corresponding to the basic region of GCN4 only (GCN4-br). Wang et al. could show that GCN4-br binds to AP-1 and ATF/CREB with similar affinity and a stoichiometry of 2:1. The ΔG of binding of GCN4-br to DNA is approximately 1.5 times lower than the ΔG of binding of C62GCN4 to DNA. This loss of affinity is due to a large loss of favorable enthalpy upon binding. This is partly compensated by the entropic penalty of binding being smaller when the monomeric construct is used. In contrast to what is observed in dimer binding the thermodynamic parameters of GCN4-br binding to AP-1 and ATF/CREB are thermodynamically very similar (*Wang et al., 2003*).

4.5.3. Kinetic Characterization of DNA Binding of C62GCN4

It is often thought that bZIP transcription factors like GCN4 have to bind to DNA as dimers because they dimerize in the absence of DNA and bind to recognition sites of dyad symmetry as dimers. This notion has become challenged by a number of thermodynamic analyses (*Hollenbeck and Oakley, 2000; Stanojevic and Verdine, 1995; Wang et al., 2003*). Kinetic measurements on GCN4 and also on Jun and Fos confirm the importance of binding through the monomer pathway, where monomeric transcription factors bind to DNA and dimerization happens on the DNA. Such a scenario helps to avoid the bottleneck that can occur when the dimerization becomes the rate-limiting step, namely when the rate of binding to DNA is higher than the rate of dimerization. Kinetic experiments conducted by Berger et al. in 1998 show that DNA binding of C62GCN4 does in fact occur at a much higher rate than dimerization. Significantly the diffusion limited association rate ($> 10^8$ M s^{-1}) does not change when using GCN4 derivatives that cannot form dimers at all or derivatives that are permanently dimerized by introduction of a disulfide bond at the C-terminus of the peptides (Berger 1998). Clearly the monomer-dimer equilibrium does not influence the overall rate of DNA recognition and binding, indicating that GCN4 can bind to ATF/CREB either as a dimer or as a monomer.

Fos and Jun are bZIP transcription factors that are structurally similar to GCN4 and bind the AP-1 site that is also the target site for GCN4 (*Hurst, 1996*). Unlike GCN4 that homodimerizes, Fos and Jun have to form a heterodimer called transcription factor AP-1 to bind to their recognition site in a stable manner. The formation of the Fos·Jun·AP-1 complex was analyzed kinetically by labeling Fos and Jun, respectively, with fluorescent dyes that constitute a FRET (fluorescence resonance energy transfer) pair. The experiments show that

the formation of Fos·Jun·AP-1 also occurs through the monomer pathway, in spite of the Jun Fos heterodimer being more stable than the GCN4 homodimer (*Kohler and Schepartz, 2001*).

4.5.4. Mutagenesis Studies of the Basic Region of GCN4

A number of comprehensive reports regarding the effect of mutations in the basic region of GCN4 on the DNA binding affinity and specificity have been published (*Kim et al., 1993; Suckow et al., 1994; Suckow et al., 1994; Suckow et al., 1993; Suckow et al., 1993*). Structural studies show that residues 232 to 249 are involved in DNA binding and direct base contacts are mediated by position 235-243. Sequence alignments show that the two most highly conserved positions, N235 and R243 and the three moderately conserved positions, A238, A239 and S242 are found among those residues that contact the bases directly. In mutational analysis changing in the most conserved amino acid residues often abolishes binding; however, their replacement can also lead to new DNA-binding specificities and changes in preferred half-site spacing (*Kim et al., 1993; Suckow et al., 1994*). Similarly changes in the moderately conserved residues A238, A239 and S242 can introduce changes in the specificities and half-site spacing requirements, binding to cognate sequences is less frequently abolished than when N235 and R243 are changed (*Kim et al., 1993; Suckow et al., 1993*).

Nevertheless, the way the basic regions of leucine zippers recognize their target sequences and distinguish them from other similar sequences and the role of the non-conserved residues in this are not entirely clear. One non-conserved residue in GCN4, T236, controls specificity by positioning the residues that directly contact the DNA, T236 itself contacts only the backbone (*Suckow et al., 1993*). The residues that make contacts to the rims of the bases in the major grooves are largely conserved among bZIP transcription factors; even those that recognize different target sites. These direct contacts between bases and amino acid residues thus seem to be only one of the mechanisms that confer binding specificity. Another mechanism thought to influence binding selectivity is the recognition of small structural deviations of DNA conformation that depend on the DNA sequence by the highly variable residues (*Hurst, 1996*). In order for the residues of the basic region to be appropriately positioned to recognize DNA the basic region has to undergo a structural reorganization from what appears to be a random coil to an α-helix (*Hollenbeck et al., 2002*). The entropic cost that is incurred by this can lead to a reduction of specificity if the protein could bind to non-specific sites without structural reorganization. This effect can be minimized when the

entropic penalty of helix formation is reduced by increasing the helical propensity of the basic region. Mutants were created that have increased helical propensity of the basic region through changing residues that are located N-terminally of the basic region and do not contact the DNA and the binding specificity of these mutants was measured. The results indicate that the helix propensity influences the binding affinity only in constructs in which the dimerization region is removed and replaced by a disulfide bridge. When the entire leucine zipper and the basic region is present the binding affinity and specificity is unchanged by the helix stabilization (*Hollenbeck et al., 2002*). This observation highlights a typical problem in the characterization of binding affinity and specificity not only of bZIP transcription factors but of virtually all protein DNA interactions. Frequently experiments are carried out at non-physiological temperatures, in the absence of substantial crowding, without competitor DNA or competitor protein, with small peptides that retain the ability to bind to DNA specifically, rather than the entire transcription factor and with small pieces of DNA instead of the DNA found in cells which usually many millions of base pairs long. The results obtained in this way may be of limited relevance for the function and behavior of a transcription factor in the cell nucleus. An insightful study by Suckow et al. compares the binding of various mutants of a GCN4 peptide consisting of leucine zipper and basic region to different target sequences in-vitro with in-vivo activation of transcription by full-length GCN4 (*Suckow and Hollenberg, 1998*). They find that in-vitro binding of a GCN4 peptide to a 52 bp DNA containing a binding site does not predict in-vivo transcription activation by corresponding full-length GCN4 of a gene under the control of a promoter corresponding to the binding site used for the in-vitro assay. Especially notable was the failure of full length GCN4 to activate from a palindromic ATF/CREB promoter, presumably the transcription factor cannot induce the bend in the DNA, which is necessary for binding to an ATF/CREB sequence when it is part of a chromosome and subject to genomic constraints. Another possibility is that the transcription factor can bind to a palindromic recognition sequence but without bending the DNA and therefore is only able to contact one of the half-sites forming a complex from which transcription initiation does not take place. However, changes of specificity upon mutations in the basic region involving the DNA half-sites themselves, rather than the half-site spacing often agree qualitatively in the in-vitro and in-vivo assay suggesting that in-vitro studies are not totally irrelevant biologically.

4.6. Objectives of the Presented Binding Study

The results of the investigations into DNA binding properties of GCN4 dimers and GCN4 monomers and their comparison raise a further question: Does GCN4 bind to its target DNA as a monomer or as a dimer or do both of these pathways occur depending on the circumstances? These questions are the focus of the second part of this thesis and can be answered by designing experiments in which various steps of the process that lead to the ternary complex of a dimeric transcription factor and DNA can be followed in isolation.

4.6.1. Thermodynamic Cycle of DNA Binding and Dimerization of C62GCN4

Figure 4.5 illustrates the two different pathways by which GCN4 can bind DNA. In the dimer pathway ($K_1 \times K_2$) two dissociation constants can be obtained: the dissociation constant of two GCN4 monomers that form a dimer (K_1) and the dissociation constant of the ternary complex formed by this dimer and the DNA (K_2). Correspondingly, there are four rates constants: the rate constants of dimer association and dissociation and the rates constants of association and dissociation of the DNA and the dimeric GCN4. Similarly the monomer pathway of binding ($K_3 \times K_4$) can be described by two dissociation constants: The two dissociation constants are for the binding of the first (K_3) and second monomer (K_4) to DNA, respectively. The four rate constants are the association and dissociation rate constants of the first and second monomer with DNA, respectively. Of course, binding can proceed through the monomer and the dimer pathway simultaneously giving rise to four different dissociation constants and eight different rate constants. In this case, the importance of the two pathways may vary depending on the starting conditions. This last arrangement constitutes a thermodynamic cycle in which the monomer and the dimer pathway of binding are thermodynamically equivalent ($K_1 \times K_2 = K_3 \times K_4$). It follows that energy is neither lost nor gained as the molecules associate and dissociate under equilibrium conditions.

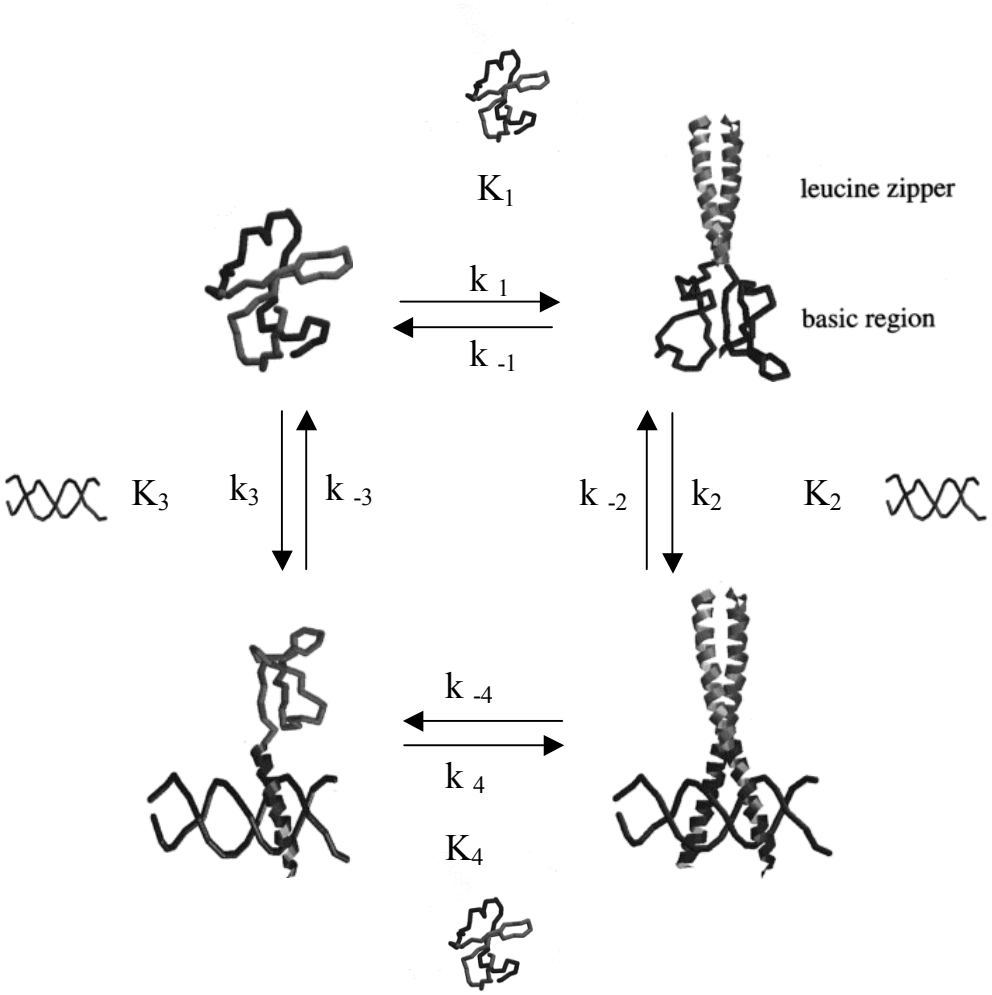


Figure 4.5 Illustration of the dimer and the monomer pathways of GCN4 binding to a cognate DNA. Because the pathways are energetically equivalent they constitute a thermodynamic cycle where $K_1K_2 = K_3K_4$. Figure adapted from (*Berger et al., 1998; Cranz et al., 2004*).

4.6.2. Experimental Investigation of a Thermodynamic Cycle

The study presented here uses various derivatives of C62GCN4 to follow the processes that lead to the ternary complex of a dimeric transcription factor and DNA in isolation. Firstly, the stability and the association and dissociation rate constants of the C62GCN4 dimer are already known. Additionally, the binding and dissociation rate constants of the dimer to the cognate DNA can be measured either by working at protein concentrations where C62GCN4 is more than 80% dimeric or by using derivatives of C62GCN4 that are covalently dimerized by the introduction of a C-terminal extension containing cysteine, which forms intermolecular disulfide bridges. The dissociation rate constant of a dimer from the DNA recognition site can also be measured by a displacement experiment with a complex consisting of the DNA and a covalently dimerized GCN4 dimer as described in "Measuring Kinetic Parameters for Protein-DNA Interaction".

The binding and dissociation rate constants of monomeric GCN4 and DNA can be measured by using GCN4 derivatives that do not have a leucine zipper region or by using derivatives where monomeric C62GCN4 is prevented from acquiring another DNA-binding element by dimerization because it is covalently linked to a leucine zipper domain without a basic region. This derivative we call "one-legged" GCN4. Additionally, the dissociation rate constant of monomeric GCN4 can be measured by a displacement experiment with this complex as described in "Measuring Kinetic Parameters for Protein-DNA Interaction".

If equilibrium dissociation constants can be assigned to three steps of a thermodynamic cycle, the dissociation constant of the reaction described by the fourth dissociation constant can be calculated. In the thermodynamic cycle described here the fourth step is the binding of monomeric GCN4 to a DNA recognition sequence, in which one half-site of the recognition sequence is already occupied by another monomer. The respective dissociation rate constant is the rate of dissociation of a GCN4 monomer from a complex in which dimeric GCN4 is bound to DNA.

4.6.3. Main Findings of the GCN4 Binding Study

In the presented study, the equilibrium dissociation constants and the association and dissociation rate constants of monomeric GCN4 and of dimeric GCN4 to a DNA containing the palindromic CRE binding site were measured directly. It was found that the rate of association of DNA with monomeric and dimeric transcription factor is similar, 5×10^8 $M^{-1}s^{-1}$ and 3×10^8 $M^{-1}s^{-1}$, respectively. Note that dimeric GCN4, but not monomeric GCN4, binds DNA through an intermediate. The rate of interconversion of the intermediate to the final complex was estimated to be ~10 s^{-1}, and the rate constant of conversion of the final complex to the intermediate was estimated to be 0.1 s^{-1}. Monomeric transcription factor dissociates with a rate of ~50 s^{-1} while dimeric transcription factor dissociates with a rate of ~30 s^{-1}. The dissociation constant of dimeric transcription factor is thus 1nM ([30 $s^{-1}\times0.1$ s^{-1}] / [3×10^8 $M^{-1}s^{-1}\times10$ s^{-1}]) and 100-fold smaller than that of monomeric GCN4.

The association of a second monomer of GCN4 with a complex of monomeric GCN4 and the DNA with the CRE binding site cannot be measured directly. However, because the equilibrium dissociation constants of the other three reactions in the thermodynamic cycle (Figure 4.5) are known, the equilibrium dissociation constant for binding of the second

monomer to the CRE site could be calculated as 6×10^{-11} M^{-1}. We assumed that the association rate constant of a second monomer to the DNA is the same as the association rate constant of a first monomer, 5×10^{8} $M^{-1}s^{-1}$, and thus concluded that the dissociation rate constant of a second monomer is 0.03 s^{-1}. From this data the degree of cooperativity of DNA binding can be quantified as the ratio between the dissociation rate constants of the first and the second monomer; cooperativity is ~1600.

Because all of the rate constants describing the thermodynamic cycle in Figure 4.5 could be either measured directly or were estimated, the time course of complex formation through the monomer and dimer pathway could be calculated using numerical integration. When the binding reaction in the calculation is started from 100 nM GCN4 monomer and 5 nM DNA the binding occurs primarily through the monomer pathway. However, when the 100 nM of GCN4 are allowed to reach monomer-dimer equilibrium, DNA binding occurs primarily through the dimer pathway. Thus, the importance of the monomer and the dimer binding pathway of GCN4 may primarily depend on the GCN4 concentration in the cell nucleus.

5. PUBLISHED MANUSCRIPT: "MONOMERIC AND DIMERIC bZIP TRANSCRIPTION FACTOR GCN4 BIND AT THE SAME RATE TO THEIR TARGET DNA-SITE"

Monomeric and Dimeric bZIP Transcription Factor GCN4 Bind at the Same Rate to Their Target DNA Site†

Susanne Cranz, Christine Berger, Antonio Baici, Ilian Jelesarov, and Hans Rudolf Bosshard*

Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received September 3, 2003

ABSTRACT: Basic leucine zipper (bZIP) transcription factors are dimeric proteins that recognize dyadic and mostly palindromic DNA sites. Dimerization of bZIP transcription factor GCN4 is linked to the folding of its C-terminal leucine zipper domain. However, monomeric GCN4, lacking a folded leucine zipper, also recognizes the DNA site with dimerization taking place on the DNA. Here we report the kinetics of DNA recognition by unfolded monomeric and folded dimeric derivatives of GCN4 using a 19 bp double-stranded DNA containing a palindromic CRE site. The rate of DNA binding of both monomeric and dimeric GCN4 has a bimolecular rate constant of $3-5 \times 10^8\ M^{-1}\ s^{-1}$, which is near the diffusion limit. Because the rate of dimerization of GCN4 is slower ($1.7 \times 10^7\ M^{-1}\ s^{-1}$) than the rate of DNA association, the formation of the dimeric GCN4−DNA complex through consecutive binding of two monomers (monomer pathway) is faster when starting from free monomers. Thus, the results presented here support facilitated and rapid target recognition by the monomeric transcription factor. However, DNA binding of preformed folded dimeric GCN4 is as rapid as complex formation through the monomer pathway. Therefore, the monomer and dimer pathways are kinetically equivalent if monomeric and dimeric GCN4 are at equilibrium. Hence, the dimer pathway may also have a role under *in vivo* conditions. The lower affinity of GCN4 in which two DNA contacting residues have been mutated is due exclusively to the faster dissociation of the mutant protein−DNA complex and not to slower complex formation.

Basic leucine zipper (bZIP)[1] transcription factors make up a large family of DNA binding proteins involved in the regulation of DNA transcription (*1*−*4*). These factors recognize promoter and enhancer regions of transcribed genes and, together with other protein factors, contribute to the efficiency by which RNA polymerase binds and initiates transcription. In the crystal structure of bZIP−DNA complexes, the dimeric protein binds to a DNA site with dyad symmetry, each monomer of the bZIP factor recognizing one half-site (*5*−*7*). In view of this structure, and since GCN4 and other bZIP factors dimerize in the absence of DNA, it has been thought that bZIP factors need to dimerize to recognize the DNA target site. However, the observed rapid rates of DNA binding could not be accounted for if formation of a dimeric bZIP peptide had to precede DNA binding (*8*−*11*). Thus, it has been proposed that monomeric transcription factors can recognize DNA and that these monomers dimerize while bound to DNA (*8*). This has been confirmed by experiment for several dimeric transcription factors (*10*−*15*). In the case of heterodimeric transcription factors such as Jun and Fos, initial binding of monomeric peptides could enable recognition of a different target site (for a recent review, see ref *16*). Furthermore, a monomer binding pathway may increase specificity and prevent the transcription factor from becoming trapped at nonspecific DNA sites (*10*−*12*, *14*, *15*).

Formation of a bZIP transcription factor−DNA complex via a monomer and a dimer pathway is shown schematically in Figure 1. Step 1 followed by step 2 is the dimer pathway, and step 3 followed by step 4 is the monomer pathway. The two pathways are energetically equivalent (thermodynamic cycle). However, there could be a kinetic preference for one or the other. Indeed, there is circumstantial evidence of a preferential monomer pathway even in the case of a transcription factor with a thermodynamically stable dimeric form (*15*).

For a direct comparison of the rates of formation of the peptide−DNA complex along the two pathways, one needs to separately study the kinetics of each step of Figure 1. Rapid DNA binding impedes quantitative determination of the DNA binding rates of dimeric (step 2) and monomeric (steps 3 and 4) GCN4 derivatives. As a result, we have previously only shown that an unfolded basic region peptide binds at a nearly diffusion-controlled rate (*11*). In the work presented here, we now have achieved the goal of assigning individual rate constants to steps 2 and 3 thanks to more sensitive rapid mixing instrumentation allowing for the use of lower reactant concentrations. Together with the rate constants of step 1 determined previously (*17*), we can approximate the rate constants for step 4 and predict the

† This work was supported in part by the Swiss National Science Foundation.

* To whom correspondence should be addressed: Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. E-mail: hrboss@bioc.unizh.ch.

[1] Abbreviations: AP-1, activation protein-1 recognition site; bZIP, basic leucine zipper; CRE, cyclic AMP response element; CRE_{19}^{F}, double-stranded 19mer oligonucleotide containing the CRE site with the fluorescence marker NBD attached to a phosphorothioate bond preceding the recognition site; GCN4, general control of amino acid synthesis non-derepressible mutant 4 (see Table 1 for GCN4 derivatives used in this study); NBD, *N*,*N*′-dimethyl-*N*-acetyl-*N*′-(7-nitrobenzo-2-oxa-1,3-diazol-4-yl)ethylenediamine.

10.1021/bi0355793 CCC: $27.50
Published on Web 12/30/2003

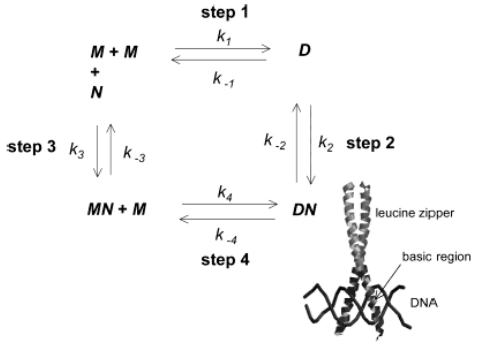


FIGURE 1: Dimer pathway (steps 1 and 2) and monomer pathway (steps 3 and 4) of association of yeast transcription factor GCN4 with its target DNA. In the dimer pathway, formation of the folded dimeric leucine zipper (step 1) precedes DNA binding of the dimer (step 2). In the monomer pathway, two monomers bind sequentially (steps 3 and 4) and the leucine zipper forms only when the second monomer binds (step 4). The reactions form a thermodynamic cycle so that $K_1K_2 = K_3K_4$ or $(k_{-1}k_{-2})/(k_1k_2) = (k_{-3}k_{-4})/(k_3k_4)$. Step 1 was studied previously (*17*). Step 2 is studied here using the derivatives C62GCN4, C62GCN4$_{SS}$, C62GCN4$_{AK}$, and BR$_{SS}$ and step 3 using derivatives C62GCN4$_{SS}^{one-leg}$ and BR (see Table 1 for structures of GCN4 derivatives). The three-dimensional structure of the GCN4−DNA complex (bottom right) is adapted from ref *7*.

overall rates of DNA binding via the monomer and dimer pathway, respectively. The results demonstrate that in the isolated system studied here, which is composed of a 19mer double-stranded DNA target and the 62-residue C-terminal DNA-binding domain of GCN4, both the monomeric and the dimeric transcription factor recognize the palindromic CRE target site at the same rapid rate. Since the rate of formation of the dimer in step 1 is significantly slower than the rates of DNA binding in steps 2−4, the monomer pathway is more rapid than the dimer pathway when starting from two monomeric GCN4 proteins and no dimer, but not when monomeric and dimeric GCN4 are at equilibrium.

EXPERIMENTAL PROCEDURES

Materials

The plasmid pET-3AC62GCN4 was used to express the 62-residue C-terminal fragment (residues 220−281) of GCN4 in *Escherichia coli* (*18*) to obtain C62GCN4 and C62GCN4$_{SS}$ as described previously (*19*). GCN4 variants BR, BR$_{SS}$, and LZ (see Table 1 for sequences) were chemically synthesized by employing the Fmoc strategy (*20*). The variant C62GCN4$_{SS}^{one-leg}$, termed "one-legged", was produced by subjecting equal amounts of reduced C62GCN4$_{SS}$ and LZ to air oxidation and purifying the products by reverse phase HPLC. Mutant C62GCN4$_{AK}$ was obtained by PCR mutagenesis according to a modified protocol (*21*). The mutations were introduced in two consecutive steps using the following templates: 5′-CAA CTT TCT CGC TTT AGA ACG CCT GCG-3′ for R243K, 5′-GGC GGC TTC AGT GGC TCT AGC ACG TTT-3′ for N235A, and 5′-ATA ATG GCC TGC TTC TCG CCG AAA CG-3′ as the back primer for both mutations. Changes with respect to the wild-type sequence are underlined.

The identity and purity of all peptides were determined by ion spray mass spectrometry. Peptide concentrations were calculated from the absorption at 280 nm using the following absorption coefficients (*22*): $\epsilon = 1280$ cm^{-1} M^{-1} for C62GCN4 and C62GCN4$_{AK}$, $\epsilon = 1340$ cm^{-1} M^{-1} for C62GCN4$_{SS}$, BR$_{SS}$, and C62GCN4$_{SS}^{one-leg}$, and $\epsilon = 1400$ cm^{-1} M^{-1} for BR. All concentrations refer to a single peptide chain, unless indicated otherwise.

The oligonucleotide TGG*t*AGATGACGTCATCTCC containing a phosphorothioate group between the third and fourth bases was obtained from MicroSynth (Balgach, Switzerland). It was labeled with the fluorescent group NBD (Molecular Probes, Eugene, OR) to obtain the fluorescent oligonucleotide CRE$_{19}^{F}$. The identity and purity of the product were confirmed by ion spray mass spectrometry. CRE$_{19}^{F}$ or CRE$_{19}$ (oligonucleotide without the fluorescent group) was annealed by heating to 90 °C and slow cooling to room temperature. The concentration of single-stranded DNA was calculated from the absorption at 260 nm using an ϵ of 198.1 mM^{-1} cm^{-1} for CRE$_{19}^{F}$ and an ϵ of 180.5 mM^{-1} cm^{-1} for CRE$_{19}$ (*23*). Given concentrations always refer to double-stranded DNA.

Methods

Buffer. All experiments were conducted at 6.5 °C in 50 mM Tris-HCl, 10 mM NaCl, and 10 mM $MgCl_2$ (pH 7.8, ionic strength of 68 mM).

Kinetic Experiments. Rates of peptide−DNA association were studied with an SX.18MV-R stopped-flow spectrometer from Applied Photophysics (Surrey, U.K.). Equal volumes of reagent solutions were mixed (dead time of <3 ms), and the change in fluorescence emission above 515 nm was measured after excitation at 495 nm. Six or more firings were averaged for each kinetic trace. In one series of experiments, equally or similarly concentrated solutions of CRE$_{19}^{F}$ and peptide were mixed. In another series, the peptide was present in excess over CRE$_{19}^{F}$, the concentration of which was kept constant at 20 nM.

To follow the dissociation of the peptide−DNA complexes, 50−100 nM CRE$_{19}^{F}$ and 50−100 nM peptide (monomer or dimer) were pre-equilibrated to form the peptide−DNA complex. The pre-equilibrated complex was mixed with nonfluorescent CRE$_{19}$ in a 40-fold excess over CRE$_{19}^{F}$. The fluorescence decrease caused by the exchange of CRE$_{19}$ for CRE$_{19}^{F}$ was followed with the fluorescence emission above 515 nm (excitation at 495 nm). When the dissociation reaction was slow, mixing was performed manually and the fluorescence decrease at 544 nm (excitation at 495 nm) was followed with a Spex Fluorolog instrument. In manual mixing (mixing time of 5−10 s), 50 μL of CRE$_{19}$ was added to 1.8 mL of a stirred solution of the preformed peptide−CRE$_{19}^{F}$ complex. We did not correct for the small dilution of the preformed complex on addition of CRE$_{19}$.

Equilibrium Fluorescence Titration and Isothermal Titration Calorimetry. Fluorescent titrations were performed in a Perkin-Elmer LS50 luminescence spectrometer by stepwise addition of small volumes of a 5−200 μM peptide solution to 1.8 mL of 125 nM CRE$_{19}^{F}$ placed in the cuvette, and the fluorescence emission in the range of 510−600 nm (excitation at 495 nm) was integrated. Peptide BR was titrated to the nonfluorescent labeled AP-1 site in a titration calorimeter (MCS ITC, MicroCal Inc., Northampton, MA) as described previously (*19*). Calorimetric titrations were performed at 6.5 °C in 50 mM sodium phosphate buffer and 100 mM NaCl (pH 7).

Table 1: Derivatives of Yeast Transcription Factor GCN4 and the Fluorescent Target DNA Site Used in This Study[a]

	Basic region leucine zipper
C62GCN4 [b]	MIVPESSDPAAL*KRARNTEAARRSRARK*LQRMKQ-LEDKVEELLSKNYHLENEVARLKKLVGER
$C62GCN4_{SS}$	MIVPESSDPAALKRARNTEAARRSRARKLQRMKQ-LEDKVEELLSKNYHLENEVARLKKLVGERGSGC | MIVPESSDPAALKRARNTEAARRSRARKLQRMKQ-LEDKVEELLSKNYHLENEVARLKKLVGERGSGC
$C62GCN4_{AK}$ [c]	MIVPESSDPAALKRARATEAARRSKARKLQRMKQ-LEDKVEELLSKNYHLENEVARLKKLVGER
$C62GCN4_{SS}^{one\text{-}leg}$	MIVPESSDPAALKRARNTEAARRSRARKLQRMKQ-LEDKVEELLSKNYHLENEVARLKKLVGERGSGC | Ac-RMKQ-LEDKVEELLSKNYHLENEVARLKKLVGERGSGCG-NH_2
LZ	Ac-RMKQ-LEDKVEELLSKNYHLENEVARLKKLVGERGSGCG-NH_2
BR	Ac-YPESSDPAALKRARNTEAARRSRARKLQRMKQGGC(StBu)G-NH_2
BR_{SS}	Ac-YPESSDPAALKRARNTEAARRSRARKLQRMKQGGCG-NH_2 | Ac-YPESSDPAALKRARNTEAARRSRARKLQRMKQGGCG-NH_2
CRE_{19}^{F} [d]	5′-TGG[NBD]AGATGACGTCATCT···CC-3′ 3′-CC···TCTACTGCAGTAGA[NBD]GGT-5′

[a] Sequence differences with respect to the wild-type GCN4 sequence are underlined. Ac, N^{α}-acetyl; NH_2, amide; StBu, thio-*tert*-butyl. [b] This sequence corresponds to C-terminal residues 220−281 of GCN4 and the N-terminal Met from the expression construct (*18*). Residues of the basic region contacting the DNA are italicized. [c] Basic region N235A and R243K mutations (*32*) are underlined. [d] CRE site with the NBD fluorescence tag, with the palindromic recognition sequence underlined.

Data Analysis. A sum of exponentials was used to fit the kinetic traces according to

$$F(t) = F_0 + \sum_i a_i[1 - \exp(-k_i^{app}t)] \quad (1)$$

where k_i^{app} values are apparent rate constants, a_i values are amplitudes, and F_0 is the fluorescence signal at time zero. Kinetic traces for monomer binding (step 3 of Figure 1) were best fit by a single-exponential fluorescence increase, and rate constants were obtained from (*24*)

$$k_3^{app} = k_3([M_{eq}] + [N_{eq}]) + k_{-3} \quad (2)$$

where $[M_{eq}] + [N_{eq}]$, the sum of the concentrations of the peptide monomer and single-stranded DNA at equilibrium, was estimated using the K_3 determined by fluorescence titration. When the total peptide concentration was much larger than the total DNA concentration, $[M_{eq}] + [N_{eq}]$ was approximated by $[M_{tot}]$. Alternatively, k_3 and k_{-3} of single-exponential kinetic traces were obtained by solving eq 3:

$$\frac{d[MN]}{dt} = k_3([N_0] - [N])([M_0] - [M]) - k_{-3}[MN] \quad (3)$$

where $[M_0]$ and $[N_0]$ refer to the initial peptide and DNA concentration, respectively. At any time point of the reaction

$$[MN(t)] = [N_0] - [N(t)] \quad (4)$$

where $[MN(t)]$ is the integral of eq 3. The decrease of free DNA with time, $[N(t)]$, is related to the fluorescence increase described by the kinetic trace according to

$$\frac{[N(t)]}{[N_0]} = \frac{F_\infty - F(t)}{F_\infty - F_0} \quad (5)$$

where $F(t)$ is the fluorescence signal at time t and F_0 and F_∞ are the fluorescence signals at time zero and infinite time, respectively. Combining eqs 4 and 5 gives

$$F(t) = F_\infty - \frac{(F_\infty - F_0)([N_0] - [MN(t)])}{[N_0]} \quad (6)$$

To obtain k_3, k_{-3}, F_0, and F_∞, kinetic traces of step 3 were analyzed by nonlinear fitting with the help of eq 6 using the integral of eq 3 to calculate $[MN(t)]$. Since simultaneous determination of all four parameters tended to produce poor estimates of k_3 and k_{-3} because of ill-defined arrays during the fitting procedure, F_∞ was estimated from the experimental trace and kept constant during the nonlinear fitting procedure, which was performed with the program SigmaPlot (version 7.0).

The best fit for dimer binding (step 2 of Figure 1) was obtained by a two-exponential fluorescence increase. The rate constants were obtained from (*24*)

$$k_2^{app} = k_2([D_{eq}] + [N_{eq}]) + k_{-2} \quad (7a)$$

$$k'^{app}_2 = k'_2\frac{k_2([D_{eq}] + [N_{eq}])}{k_2([D_{eq}] + [N_{eq}]) + k_{-2}} + k'_{-2} \quad (7b)$$

The above equations describe a biphasic reaction between two reactants in which the rapid phase exhibits a linear concentration dependence and the slow phase a hyperbolic dependence (*24*). k_2^{app} is the apparent rate constant of the first rapid phase of the experimental trace and k'^{app}_2 that of the slow phase. $[D_{eq}] + [N_{eq}]$ was estimated with the help of K_2 from fluorescence titration. Under the given experimental conditions, $k_2([D_{eq}] + [N_{eq}])$ was much larger than k_{-2}, so eq 7b can be reduced to $k'^{app}_2 \cong k'_2 + k'_{-2}$.

Dissociation constants, K, were obtained from fluorescence titration data by fitting the integrated fluorescence signal, F, according to

$$F = [(F_\infty - F_0)([N_0] + [P_0] + K - \sqrt{([N_0] + [P_0] + K)^2 - 4[N_0][P_0]}]/(2[N_0] + F_0) \quad (8)$$

where F_0 and F_∞ are the initial and final fluorescence signal, respectively. In the case of the monomeric derivatives BR and C62GCN4$_{SS}^{one-leg}$, $[P_0]$ is the total concentration of a single DNA binding site and $[N_0]$ the total concentration of one half-site of double-stranded CRE$_{19}^{F}$. In the case of dimeric derivatives C62GCN4, C62GCN4$_{SS}$, C62GCN4$_{AK}$, and BR$_{SS}$, $[P_0]$ is the total concentration of the dimeric peptide and $[N_0]$ the total concentration of double-stranded CRE$_{19}^{F}$.

Error Estimates. Standard errors of the regression of the fits to the experimental data are shown unless indicated.

RESULTS

Design of Monomeric and Dimeric Derivatives of GCN4 for Analyzing Single Steps of the Monomer and Dimer Pathway. Transcriptional activator GCN4 is a homodimeric protein composed of two polypeptide chains of 281 residues each (*25*, *26*). The 30 C-terminal residues contain four heptadic repeats that induce dimerization by formation of a leucine zipper (*5*, *27*). Upon DNA binding, the α-helices of the leucine zipper extend N-terminally into a long bipartite structure that grips the large groove of double-stranded DNA (*1*). The structure is shown in outline at the bottom right of Figure 1. The C-terminal 60-residue segment of GCN4 is sufficient for DNA recognition (*28*). It is composed of a basic region responsible for specific DNA recognition, followed by the C-terminal leucine zipper (Table 1). The basic region recognizes the target sequences ATGACTCAT named the AP-1 site and ATGACGTCAT named the CRE site (*29*). The latter differs from the AP-1 site by an additional G (underlined), which makes the CRE site palindromic. Both target sites are recognized with the same overall affinity (*19*, *30*). The 62 C-terminal residues of GCN4 (residues 220−281), called C62GCN4 (*19*), together with a 19 bp DNA comprising the CRE site were used in this study. DNA association was monitored with the help of the fluorescently labeled oligonucleotide CRE$_{19}^{F}$ (Table 1). The level of fluorescence emission of the NBD group of CRE$_{19}^{F}$ increases on binding to the basic region of GCN4, the fluorescence change being specific for the AP-1 and CRE sites (*11*). CRE$_{19}^{F}$ harboring the CRE site was used here because it exhibits a somewhat larger fluorescence change upon binding than the NBD-labeled AP-1 site (*11*).

The combined monomer and dimer pathways are described by a minimum of four equilibrium and eight rate constants (Figure 1). In principle, all the rate constants may be obtained by simultaneous fitting of kinetic traces acquired by mixing the peptide and DNA at a series of different concentrations using a numerical fitting algorithm (*31*). In practice, however, experimental noise prevents the unequivocal assignment of rate constants to each step (*15*). As an alternative, we decided to study single steps of Figure 1 with the help of appropriate monomeric and dimeric derivatives of GCN4 (Table 1).

Thermodynamics and kinetics of reaction step 1 (leucine zipper formation in the absence of DNA) have been analyzed previously (*17*). Association of dimeric GCN4 with DNA in reaction step 2 was studied in four different ways: (i) with the disulfide-linked, permanently dimeric derivative C62GCN4$_{SS}$, (ii) with C62GCN4 sufficiently concentrated to make the monomer concentration negligible, (iii) with disulfide-linked basic region peptide BR$_{SS}$ lacking the leucine zipper domain, and (iv) with sufficiently concentrated mutant derivative C62GCN4$_{AK}$ in which two conserved residues contacting the DNA are mutated to weaken DNA binding without destabilizing the leucine zipper dimer (*32*). Reaction step 3, which is DNA association of the monomer, was investigated with two derivatives: the basic region peptide BR and the one-legged derivative C62GCN4$_{SS}^{one-leg}$. The latter is the disulfide-linked dimeric leucine zipper with only a single basic DNA-binding region. Finally, estimates of equilibrium and rate constants for the binding of the second monomer in step 4 were obtained from the thermodynamic cycle described in Figure 1.

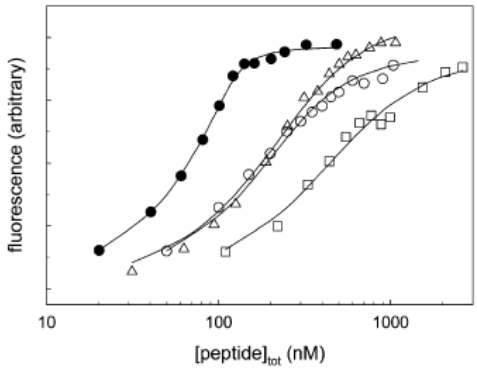


FIGURE 2: Titration of 125 nM double-stranded CRE$_{19}^{F}$ with C62GCN4 (●), C62GCN4$_{AK}$ (△), C62GCN4$_{SS}^{one-leg}$ (○), and BR (□). Solid lines are best fits according to eq 8 describing a 1/1 complex.

Equilibrium Fluorescence Titration. Dissociation constants of the peptide−DNA complexes were calculated from the fluorescence increase observed when CRE$_{19}^{F}$ was titrated with the monomeric or dimeric peptide. Representative titrations are shown in Figure 2, and constants K_2 and K_3 are given in Tables 2 and 3, respectively. As expected, the best binders are the dimeric wild-type peptides C62GCN4 and C62GCN4$_{SS}$. Their K_2 values (Table 2) are around 2 nM, the error of K_2 being large since fluorescence titrations had to be performed at peptide and DNA concentrations that were much higher than K_2. Independent isothermal titration calorimetry of CRE$_{19}$ with C62GCN4 at 9 °C, pH 7.6, and an ionic strength of 190 mM yields a K_2 of 6.3 ± 3.3 nM (*19*). The dimeric derivative C62GCN4$_{AK}$ in which two residues contacting DNA have been mutated shows weaker binding with a K_2 of 114 ± 48 nm.

It should be noted that K_2 refers to a 1:1 complex of the dimeric peptide and double-stranded CRE$_{19}^{F}$; i.e., the binding of the two legs of the dimeric peptide to the two half-sites of CRE is regarded as a single binding event. This simplification is justified under equilibrium conditions since the dimeric peptides were either disulfide-linked (C62GCN4$_{SS}$ and BR$_{SS}$) or present at a sufficiently high concentration to make the monomer concentration negligible (C62GCN4 and C62GCN4$_{AK}$).

The value of K_3 for monomer binding was attained using monomeric BR and C62GCN4$_{SS}^{one-leg}$ as GCN4 derivatives which make contact with only one half-site. Monomeric BR is the most weakly binding derivative with a K_3 of ~400

Table 2: Kinetic Rate Constants and Equilibrium Dissociation Constants for DNA Binding of Dimeric GCN4 Derivatives (Step 2 of Figure 1)

	$D+N \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} DN^* \underset{k'_{-2}}{\overset{k'_2}{\rightleftarrows}} DN$			DNA displacement [c]	$K_2 = \frac{[D][N]}{[DN^*]+[DN]}$ [d]
	k_2 ($M^{-1}s^{-1}$) [a]	k_{-2} (s^{-1}) [a]	$k'_2 + k'_{-2}$ (s^{-1}) [b]	k_{disp} (s^{-1})	(nM)
C62GCN4	$(3.0 \pm 1.3) \times 10^8$	36 ± 15	10 ± 8	$(2.4 \pm 0.8) \times 10^{-4}$	≈ 2
$C62GCN4_{SS}$	$(3.4 \pm 1.4) \times 10^8$	25 ± 20	10 ± 8	$(5.4 \pm 1.6) \times 10^{-4}$	1.5 ± 1.0 [e]
$C62GCN4_{AK}$	$(7.7 \pm 2.5) \times 10^8$	[f]	1.1 ± 0.2	5.0 ± 0.2	114 ± 48
BR_{SS}	$(7.0 \pm 2.5) \times 10^8$	[f]	8 ± 5	≈ 0.02 [g]	58 ± 22

[a] From plot according to eq 7a. [b] From plot according to eq 7b and the limiting case of $k_2([D_{eq}] + [N_{eq}]) \gg k_{-1}$. [c] Dilution of the complex with an excess of nonfluorescent CRE_{19}. [d] Fluorescence titration of CRE_{19}^F with peptide. [e] From ref *11*. [f] Could not be determined because of the large error. [g] The displacement reaction shows two exponential phases; the overall $t_{1/2}$ is 3.5 s.

Table 3: Kinetic Rate Constants and Equilibrium Dissociation Constants for DNA Binding of Monomeric GCN4 Derivatives (Step 3 of Figure 1)

	$M+N \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} MN$		$K_3 = \frac{k_{-3}}{k_3}$	DNA displacement [c]	$K_3 = \frac{[M][N]}{[MN]}$ [d]
	k_3 ($M^{-1}s^{-1}$)	k_{-3} (s^{-1})	(nM)	k_{disp} (s^{-1})	(nM)
$C62GCN4_{SS}^{one-leg}$	$(4.6 \pm 1.2) \times 10^8$ [a]	66 ± 18 [a]	140 ± 54	6.2 ± 0.2	121 ± 11
	$(6.4 \pm 4.3) \times 10^8$ [b]	33 ± 16 [b]	52 ± 42		
BR	$(5.1 \pm 1.5) \times 10^8$ [b]	83 ± 20 [b]	163 ± 62	73 ± 5	406 ± 32

[a] From plot according to eq 2. [b] From data analysis according to eq 6. [c] Dilution of the complex with an excess of nonfluorescent CRE_{19}. [d] From fluorescence titration of CRE_{19}^F with peptide.

nM. $C62GCN4_{SS}^{one-leg}$ has a K_3 of 121 ± 11 nM, similar to the K_2 measured for the dimeric derivative $C62GNC4_{AK}$ in which two residues contacting DNA have been mutated and similar to the K_2 of the "minimal" dimer BR_{SS} composed of two disulfide-linked basic region peptides.

DNA Association of Dimeric Derivatives of GCN4. Kinetic traces for GCN4 derivatives with two DNA binding sites exhibit two exponential phases: an initial rapid phase, which depends linearly on peptide concentration, and a slow and apparently concentration-independent phase. This behavior can be explained by the following mechanism:

$$D + N \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} DN^* \underset{k'_{-2}}{\overset{k'_2}{\rightleftarrows}} DN$$

where D is the dimeric peptide, N is CRE_{19}^F, DN* is an intermediate, and DN is the final complex. The four rate constants are related to the two apparent rate constants k_2^{app} and k'^{app}_2 as described by eqs 7a and 7b in Methods. The amplitude of the slower phase described by k'^{app}_2 decreases at increasing peptide concentrations so that k'^{app}_2 could no longer be measured above ~200 nM peptide. A small amount of monomeric C62GCN4 or $C62GCN4_{AK}$ at equilibrium with the corresponding dimer could be neglected since K_1 of the monomer−dimer equilibrium (step 1 of Figure 1) is 6 nM at 5 °C (*17*), similar to the K of 8 nM reported for the leucine zipper domain alone (*33*). Therefore, C62GCN4 was 85% dimeric already at the lowest peptide concentration used in kinetic experiments. Figure 3A shows a kinetic trace for the reaction of C62GCN4 with CRE_{19}^F. Plots of the observed rate constants k_2^{app} and k'^{app}_2 versus the equilibrium concentrations ($[D_{eq}] + [N_{eq}]$) are shown in Figure 3B. While k_2 could be calculated with good accuracy, k_{-2} extrapolated from y-intercepts had a very large error. In principle, k'_2 and k'_{-2} of the slow phase can be obtained from plots according to eq 7b provided the experiment can be performed at a sufficiently low reactant concentration. This was not possible, and experimental errors were large because of the small amplitude of the second phase. Therefore, only a rough estimate of $(k'_2 + k'_{-2})$ could be obtained. Rate constants for the dimeric derivatives are summarized in Table 2.

DNA Association of Monomeric Derivatives of GCN4. Figure 4A shows a representative trace of the reaction of $C62GCN4_{SS}^{one-leg}$ with CRE_{19}^F. Binding of one-legged GCN4 and binding of the monomeric DNA binding domain BR exhibit a single-exponential phase described by

$$M + N \underset{k_3}{\overset{k_{-3}}{\rightleftarrows}} MN$$

Here, M corresponds to the single DNA-binding site of $C62GCN4_{SS}^{one-leg}$ or to a single basic region peptide BR, and N is a half-site of double-stranded CRE_{19}^F. Rate constants k_3 and k_{-3} were obtained by plotting apparent rate constants according to eq 2. The plot for $C62GCN4_{SS}^{one-leg}$ is shown in Figure 4B. Alternatively, kinetic traces were also fit according to eq 6, which is based on the integrated rate equation, eq 3 (solid line in Figure 4A). The bimolecular association rates of $C62GCN4_{SS}^{one-leg}$ and BR are $\sim 5 \times 10^8$ M^{-1} s^{-1}, indistinguishable within experimental error (Table 3). Also, the dissociation rate constants k_{-3} of the MN complexes are similar but have a large error from extrapolation to the y-axis intercept, as has k_{-3} from fitting with eq 6.

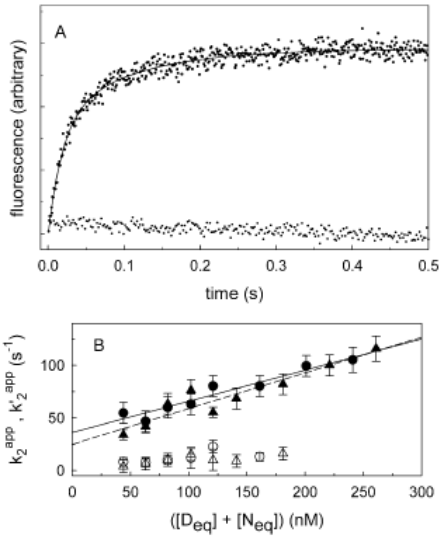


FIGURE 3: Representative kinetic trace of the association of dimeric C62GCN4 with DNA. (A) Reaction of 140 nM dimeric C62GCN4 with 20 nM $CRE_{19}{}^{F}$. The solid line is a best fit for a k_2^{app} of 53 s^{-1} (amplitude of 73%) and a $k'_2{}^{app}$ of 9 s^{-1} (amplitude of 27%). The bottom trace shows the control reaction of $CRE_{19}{}^{F}$ with the basic peptide EYQALKKKVAQLKAKNQALKKKVAQLKHKG, whose sequence is unrelated to the basic region of C62GCN4 (see ref *11* for details). (B) Apparent rate constants k_2^{app} (filled symbols) and $k'_2{}^{app}$ (empty symbols) for DNA binding of C62GCN4 (circles, solid line) and $C62GCN4_{SS}$ (triangles, dashed line) plotted according to eq 7a and 7b, respectively. Error bars indicate the data range of two or three independent experiments.

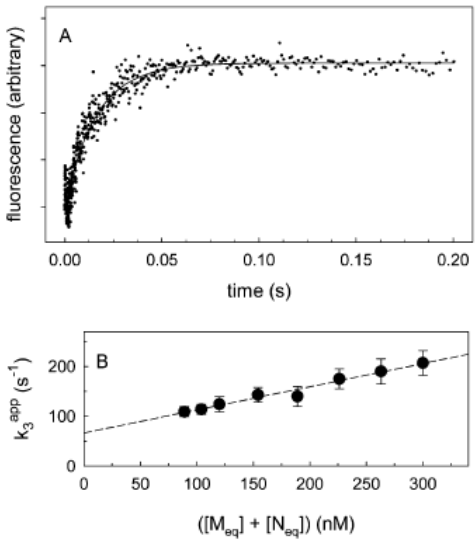


FIGURE 4: Representative kinetic trace of association of monomeric GCN4 with DNA. (A) Reaction of 35 nM one-legged derivative $C62GCN4_{SS}{}^{one-leg}$ with 35 nM $CRE_{19}{}^{F}$. The solid line is a best fit according to eq 6. (B) Apparent rate constant k_3^{app} for DNA binding of $C62GCN4_{SS}{}^{one-leg}$ plotted according to eq 2. Error bars indicate the data range of two or three independent experiments.

The dissociation constant K_3 calculated as k_{-3}/k_3 is in the range of 50−140 nM for $C62GCN4_{SS}{}^{one-leg}$, in agreement with K_3 from fluorescence titration. For the monomer BR, K_3 from kinetics is 3 times smaller than from fluorescence titration.

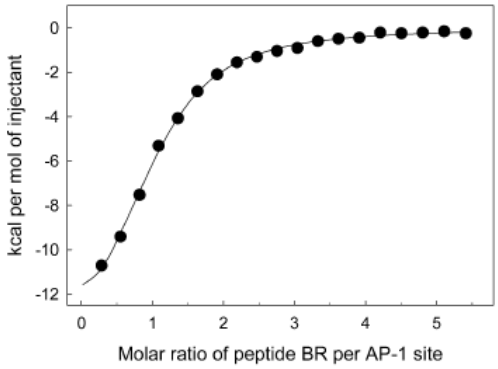


FIGURE 5: Calorimetric titration of the AP-1 site with basic region peptide BR. The solid line is a best fit for a stoichiometry n of 0.96 and a K_3 of 2 μM, which is larger than the K_3 obtained by fluorescence titration (Table 3), probably because of the higher ionic strength applied in calorimetry.

The measured rate constants and dissociation constants might not pertain to step 3 alone but could encompass steps 3 and 4 together if two monomeric peptides bind to the dyadic CRE site. This is very unlikely for $C62GCN4_{SS}{}^{one-leg}$ since the two coiled coils from two molecules of $C62GCN4_{SS}{}^{one-leg}$ would interfere with each other in the orientation seen in the crystal structure of the wild-type complex (Figure 1). Less steric hindrance is expected for the binding of two molecules of BR. Isothermal titration calorimetry showed, however, that only a single BR binds to the two half-sites (Figure 5). The stoichiometry is 0.96 mol of BR per mole of double-stranded DNA. Furthermore, the heat capacity change for binding of a single BR molecule is roughly half that for binding of dimeric C62GCN4 (I. Jelesarov and C. Berger, unpublished experiments). Hence, the kinetic and equilibrium data obtained with the monomeric derivative BR pertain to step 3 alone.

Dissociation Rates Estimated from DNA Displacement. An independent estimate of the overall dissociation rate constant of the peptide−DNA complexes was obtained from dilution of the preformed peptide−$CRE_{19}{}^{F}$ complex with an excess of nonfluorescent DNA according to

$$P{-}CRE_{19}{}^{F} + CRE_{19} \xrightarrow{k_{disp}} P{-}CRE_{19} + CRE_{19}{}^{F} \quad (9)$$

where P is either monomer M or dimer D. In the case of a complex in which only one half-site of the DNA is contacted by the peptide, the observed rate constant of DNA displacement, k_{disp}, should be equivalent to k_{-3} if the rate of rebinding of nonfluorescent CRE_{19} is much faster than the dissociation of the complex. Rebinding, calculated as k_3[CRE_{19}], was ca. 1000 s^{-1} under the chosen experimental conditions, much faster than k_{-3}. Figure 6A shows DNA displacement from the $C62GCN4_{SS}{}^{one-leg}$−$CRE_{19}{}^{F}$ complex. The reaction obeys a single-exponential decay with a k_{disp} of 6.2 s^{-1}, a value 5−10 times slower than k_{-3} obtained from binding traces (Table 3). In the case of the BR−$CRE_{19}{}^{F}$ complex, DNA displacement has a rate of 73 ± 5 s^{-1}, in good agreement with k_{-3} from binding traces (Table 3).

DNA displacement from dimeric C62GCN4 and $C62GCN4_{SS}$ is very much slower and has a half-time of ∼40 min (Figure 6B). Even DNA displacement from the complex with

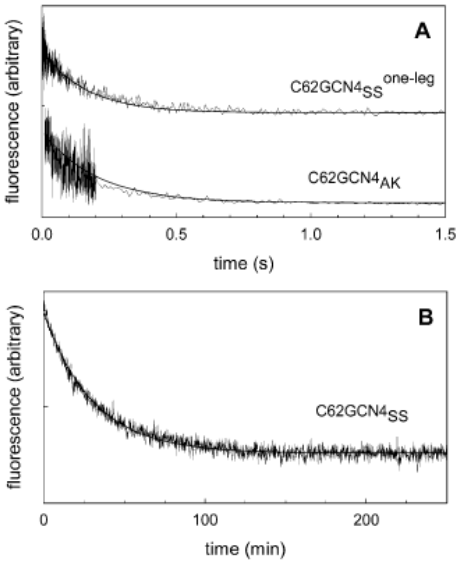


FIGURE 6: Displacement of DNA from the preformed protein−DNA complex. (A) Approximately 20 nM preformed complex composed of CRE_{19}^F and $C62GCN4_{SS}^{one-leg}$ or $C62GCN4_{AK}$ was diluted with 2 μM nonfluorescent CRE_{19}. The solid lines are best fits for k_{disp} values of 6.2 s^{-1} ($C62GCN4_{SS}^{one-leg}$) and 5 s^{-1} ($C62GCN4_{AK}$) based on the data from six independent experiments. The traces are shifted with respect to the y-axis for clarity. (B) Approximately 40 nM preformed complex composed of $C62GCN4_{SS}$ and CRE_{19}^F was diluted with 2 μM nonfluorescent CRE_{19}. The solid line is a best fit for a k_{disp} of 2.4×10^{-4} s^{-1}. Note the different scale of the time axes in panels A and B.

the "minimal" dimer BR_{SS} has a halftime of ~1 min, significantly slower than the dissociation of the one-legged complex. However, DNA displacement from the dimeric mutant $C62GNC4_{AK}$, which binds DNA more weakly than wild-type C62GCN4, is roughly as fast as DNA displacement from the monomeric one-legged peptide (Figure 6A).

DISCUSSION

Understanding the forces involved in protein−DNA recognition requires careful and systematic analysis of the thermodynamics and kinetics of protein−DNA association. Electrophoretic mobility shift assays are often used to investigate protein−DNA interactions. However, this assay may be problematic because it is not at equilibrium and complexes may dissociate during electrophoresis. Monomer−DNA complexes in particular might escape detection because they are kinetically unstable. Environmentally responsive fluorescence probes allow us to follow protein−DNA association reactions by rapid mixing in the second to millisecond time range. Patel and co-workers (*34*) and Kohler and Schepartz (*15*) attached different fluorescence tags to the leucine zipper part of Jun and Fos to monitor resonance energy transfer when the Jun−Fos heterodimer forms in the presence and absence of DNA, respectively. Via this elegant approach, it was shown that dimerization of Jun and Fos is more rapid in the presence of DNA (*15*, *34*), that the Jun−Fos−DNA complex is extremely stable (*34*), and that its formation follows the monomer pathway (*15*).

Since our goal was the direct comparison of the monomer and dimer pathway of DNA binding, following the reaction by visualizing dimer formation from a fluorescence tag in the leucine zipper domain was not applicable. The fluorescence tag was attached on the DNA instead. An NBD group was placed next to the palindromic CRE site to obtain CRE_{19}^F. Fluorescence emission of NBD reports sequence specific binding to the CRE and AP-1 sites but not unspecific, electrostatically driven DNA binding (control trace in Figure 3 and ref *11*). One disadvantage of attaching the fluorescence tag to the DNA is that experiments are restricted to short oligonucleotides since the fluorescence signal becomes very weak when long pieces of double-stranded DNA are used as these cannot be sufficiently concentrated for the kind of rapid mixing experiments performed here. In contrast, peptide binding to a heterogeneous mixture of large DNA fragments from calf thymus could be monitored when the fluorescence tag was attached to the leucine zipper domain (*15*)

Dimer Pathway. Formation of dimeric C62GCN4 in the absence of DNA (step 1 of Figure 1) has been studied in detail previously (*17*). At 5 °C, the dissociation constant of dimeric C62GCN4 is 6 nM; the rate constants of dimer formation and dissociation are 1.6×10^7 M^{-1} s^{-1} and 0.1 s^{-1}, respectively.

Binding of dimeric derivatives of GCN4 to the CRE site in a 19mer oligonucleotide exhibits two kinetic phases: a rapid, concentration-dependent phase and a slow, apparently concentration-independent phase. The reaction is interpreted as the concentration-dependent formation of an initial complex DN*, which rearranges in a concentration-independent step to the thermodynamically stable complex DN. The nature of the intermediate DN* and its rearrangement to DN are unknown. One is tempted to speculate that DN* is an intermediate with only "one leg" bound to one DNA half-site and that the rearrangement conforms to the intramolecular binding of the "second leg" to the second half-site. In support of this proposition, binding of mutant $C62GCN4_{AK}$, whose binding strength is similar to that of the monomeric derivatives, also exhibits two exponential reaction phases, whereas monomer binding is always single-exponential.

The association rate constant k_2 for the D + N → DN* reaction is 3×10^8 M^{-1} s^{-1}, which is not far from a diffusion limit of 5×10^9 M^{-1} s^{-1} estimated by the von Smolukowski relationship for the association of uncharged molecules (*35*, *36*). The diffusion limit could be larger because of electrostatic attraction between the charged reactants. Similar rates have been reported for the binding of the *E. coli* SSB protein to single-stranded DNA (*37*) and for DNA binding of the papilloma E2 protein (*38*).

The two reaction phases of dimer binding are coupled so that the DN* → DN rearrangement should exhibit hyperbolic concentration dependence (eq 7b). This was not seen because reactant concentrations had to be relatively large and the scatter of $k_2'^{app}$ was considerable. Since the amount of intermediate DN* will be small at equilibrium, k_2' has to be significantly larger than k_{-2}', which means $k_{-2}' \ll k_2'^{app}$. If we assume the equilibrium is 100 times on the side of the final complex DN, k_2' would be similar to $k_2'^{app}$ (≈10 s^{-1}) and k_{-2}' would be ~0.1 s^{-1}. Using these values of k_2' and k_{-2}' together with a k_2 of 3×10^8 M^{-1} s^{-1} and a k_{-2} of 30 s^{-1}, the overall equilibrium constant K_2 for DNA binding to

dimeric C62GCN4 and C62GCN4$_{SS}$ is 1 nM,[2] in agreement with the K_2 estimated by fluorescence titration. The overall dissociation constant of the dimer pathway is K_1K_2 and amounts to 6×10^{-18} M^2 for C62GCN4, in good agreement with a value of 4.1×10^{-18} M^2 reported for a similar dimeric GCN4 derivative bound to the CRE site (*10*).

Monomer Pathway. To study monomer binding (step 3) in isolation from dimer formation and dimer binding, we have designed the derivative C62GCN4$_{SS}^{one-leg}$. It contains the same leucine zipper domain as C62GCN4 but only a single DNA binding sequence. Because no two molecules of C62GCN4$_{SS}^{one-leg}$ can bind in the proper orientation to the CRE target site (*7*, *18*), the reaction with C62GCN4$_{SS}^{one-leg}$ conforms to step 3 alone. But even the binding of the basic region peptide BR conforms to step 3 alone, as demonstrated by isothermal titration calorimetry (Figure 5). Why no two molecules of BR can bind at the same time is not entirely clear. Since only part of BR is contacting the DNA (italicized residues in Table 1), the remaining residues of two bound molecules of BR may sterically interfere with each other unless kept properly oriented by the leucine zipper domain.

The association rate of the monomer is virtually the same as that of the dimer, 5×10^8 M^{-1} s^{-1}. The dissociation constant K_3 of step 3 is ~2 orders of magnitude larger than the K_2 of dimer binding because the complex MN dissociates faster, k_{-3} being around 50 s^{-1}. The complex with C62GCN4$^{one-leg}$ is more stable than that with BR, yet still 100 times less stable than the complex with dimeric GCN4. Kim and Little reported a K_2 of 1 μM for a complex corresponding to MN (*8*).

Step 4 of the monomer pathway could not be analyzed in a direct way. The equilibrium constant K_4 follows from the thermodynamic cycle and is 6×10^{-11} M for C62GCN4. Since the constants for the associations of monomeric and dimeric GCN4 are both $\sim 5 \times 10^8$ M^{-1} s^{-1}, it is reasonable to assume that binding of the second monomer in step 4 has a similar rate. With a k_4 of 5×10^8 M^{-1} s^{-1}, k_{-4} becomes 0.03 s^{-1}, at least 1000 times slower than the dissociation of the first monomer from the complex MN. This testifies to the high cooperativity between tethered ligands interacting with two adjacent binding sites (*39*). The degree of cooperativity can be quantified as K_3/K_4 or $(K_3)^2/(K_1K_2)$ (*8*). One obtains a value of 1600, which means that the probability of the second monomer staying on the DNA is 1600 times higher because it can dimerize with the first monomer through the leucine zipper domain.

Displacement of DNA from the Dimer−DNA Complex Is Very Slow. The rate of DNA displacement (Figure 6 and eq 9) should mirror the rate-limiting step of the dissociation of DN, which probably is the DN → DN* reaction step governed by k'_{-2}. Here, we note a discrepancy between the rate of DNA displacement and the kinetic constants for step 2. If one assumes $k'_{-2} = k_{disp} = 2.4 \times 10^{-4}$ s^{-1}, K_2 of C62GCN4 would be 3 pM,[2] 3 orders of magnitude less than K_2 from fluorescence titration. The K_2 of several bZIP transcription factor−DNA complexes is on the order of 1−10 nM, far from the picomolar range (*10*, *15*, *34*, *40*, *41*). Thus, it seems that displacement of DNA from the dimer−DNA complex does not monitor the overall dissociation rate of the complex and tends to overestimate its stability. In the presence of excess DNA during the exchange reaction, dimeric GCN4 may transiently bind to two molecules of double-stranded oligonucleotide, thereby slowing DNA exchange.

[2] Calculated from the relationship $K_2 = (k_{-2}k'_{-2})/(k_2k'_{-2} + k_2k'_2)$.

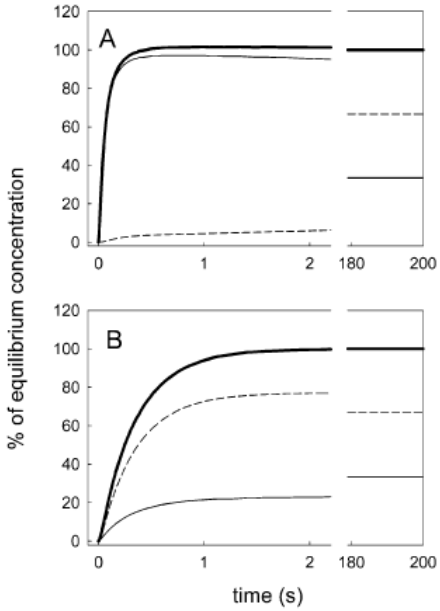


FIGURE 7: Comparison of the monomer and dimer pathways. The time course of formation of the complex DN via the monomer pathway (thin lines), the dimer pathway (dashed lines), and their sum (thick lines) are shown. (A) Forced monomer pathway. If the reaction starts from 100 nM monomer and 5 nM double-stranded DNA, more than 90% of the equilibrium concentration of the complex DN forms via the monomer pathway in less than 1 s. Although the buildup of the complex via the dimer pathway is very slow, more complex forms through the dimer pathway if the system is let to equilibrate because $k_{-1} \ll k_{-3}$. (B) *In vivo*-like situation. Letting the monomer and dimer pre-equilibrate (step 1) *before* addition of the DNA accelerates the dimer pathway. In the calculated example, 100 nM monomer was allowed to equilibrate to reach a $[M]_{equ}$ of 16 nM and a $[D]_{equ}$ of 42 nM before addition of 5 nM DNA. Calculations were performed by numerical integration using the following rate constants: $k_1 = 1.6 \times 10^7$ M^{-1} s^{-1}, $k_{-1} = 0.05$ s^{-1}, $k_2 = 3 \times 10^8$ M^{-1} s^{-1}, $k_{-2} = 30$ s^{-1}, $k'_{-2} = 10$ s^{-1}, $k'_{-2} = 0.1$ s^{-1}, $k_3 = 5 \times 10^8$ M^{-1} s^{-1}, $k_{-3} = 50$ s^{-1}, $k_4 = 5 \times 10^8$ M^{-1} s^{-1}, and $k_{-4} = 0.03$ s^{-1}.

Rates of DNA Binding through the Monomer and Dimer Pathway. We can now calculate the time course of complex formation through the monomer and dimer pathway using the experimentally determined rate constants of steps 1−3 together with the estimated rate constants of step 4. Figure 7 shows the results of a numerical integration of the four concerted reactions of Figure 1. In comparing the two pathways, one must distinguish between (i) starting the DNA binding reaction from free monomeric transcription factor alone and (ii) starting from pre-equilibrated monomeric and dimeric transcription factor in step 1. The latter condition may be closer to an *in vivo* situation in which monomeric and dimeric transcription factor compete for the DNA sites. If starting from monomeric transcription factor alone, DNA binding through the monomer pathway is very much faster than through the dimer pathway (Figure 7A). The reason is

the slower rate of monomer to dimer association in comparison to the rate of DNA binding of the monomer. However, if monomeric and dimeric transcription factor are at equilibrium before the reaction with the DNA, the dimer pathway is as rapid as the monomer pathway (Figure 7B). In the example of Figure 7B, more complex forms through the dimer pathway because [D] > [M] in pre-equilibrated step 1. It should be noted that the absolute rate of complex formation depends strongly on the initial concentrations of peptide and DNA and on the peptide/DNA ratio. For the calculation shown in Figure 7, we chose 100 nM monomer in 20-fold excess over the double-stranded DNA target. However, irrespective of the concentration and ratio of reactants, the general trend remains the same: DNA binding through the monomer pathway is more rapid if the reaction starts from monomeric transcription factor alone because dimer formation is relatively slow. However, the monomer and dimer pathways have similar rates if a substantial amount of dimer is present at the beginning of the reaction.

In their analysis of DNA binding of the Jun−Fos transcription factor based on fluorescence energy transfer between tagged Jun and Fos peptides, Kohler and Schepartz came to the conclusion that the dimer pathway is inadequate and unnecessary for complex formation (*15*). However, this conclusion holds only if all complex is formed from monomeric Jun, monomeric Fos, and DNA, which were the experimental conditions chosen by Kohler and Schepartz and in the calculations presented in Figure 7A. In a cellular environment, an equilibrium mixture of monomeric and dimeric transcription factors may be competing for DNA sites. If monomer and dimer bind at a similar rapid rate, as found here for GCN4, the choice of the pathway is influenced by the pre-existing ratio of monomeric to dimeric transcription factor. Hence, one would expect that a stable dimeric transcription factor could follow the dimer pathway under *in vivo* conditions.

We are well aware that the above conclusions are based on *in vitro* studies using a short target oligonucleotide and a truncated transcription factor. The assembly of a transcription factor−DNA complex *in vivo* is an intricate and complex affair. Not only must the transcription factor search its specific binding site in the presence of a large excess of unspecific DNA, but there also are other proteins competing for the DNA, much of it being packed in chromatin. Target finding occurs mainly by diffusion of the protein along the DNA (*42*−*44*). A monomeric transcription factor may diffuse more rapidly than a dimeric transcription factor because unspecific DNA binding of the monomer is weak. Unspecific DNA binding of the dimer could be stronger because of more nonspecific electrostatic interactions. Thus, the relative rates of diffusion of the monomer and dimer along the DNA are other variables affecting the time course of target recognition *in vivo*. Finally, accessory proteins influence the strength of the transcription factor−DNA complex (*45*). The rates of target finding and DNA binding through a monomer or dimer pathway could differ, depending on whether such accessory proteins bind to the monomeric or dimeric transcription factor, or both. In conclusion, there are many variables affecting the choice for a monomer or dimer recognition pathway *in vivo*. To clarify such complexity, future investigations must focus on the rates of diffusion of monomeric and dimeric transcription factors along the DNA and on the effect of other proteins on the stability of transcription factor−DNA complexes.

REFERENCES

1. Vinson, C. R., Sigler, P. B., and McKnight, S. L. (1989) *Science 246*, 911−916.
2. Baxevanis, A. D., and Vinson, C. R. (1993) *Curr. Opin. Genet. Dev. 3*, 278−285.
3. Hurst, H. C. (1996) *Leucine Zippers: Transcription factors*, 3rd ed., Academic Press, London.
4. Ellenberger, T. (1994) *Curr. Opin. Struct. Biol. 4*, 12−21.
5. Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) *Cell 71*, 1223−1237.
6. Glover, J. N. M., and Harrison, S. C. (1995) *Nature 373*, 257−261.
7. Keller, W., König, P., and Richmond, T. J. (1995) *J. Mol. Biol. 254*, 657−667.
8. Kim, B., and Little, J. W. (1992) *Science 255*, 203−206.
9. Park, C., Campbell, J. L., and Goddard, W. A. (1996) *J. Am. Chem. Soc. 118*, 4235−4239.
10. Metallo, S. J., and Schepartz, A. (1997) *Nat. Struct. Biol. 4*, 115−117.
11. Berger, C., Piubelli, L., Haditsch, U., and Bosshard, H. R. (1998) *FEBS Lett. 425*, 14−18.
12. Wu, X., Spiro, C., Owen, W. G., and McMurray, C. T. (1998) *J. Biol. Chem. 273*, 20820−20827.
13. Rentzeperis, D., Jonsson, T., and Sauer, R. T. (1999) *Nat. Struct. Biol. 6*, 569−573.
14. Kohler, J. J., Metallo, S. J., Schneider, T. L., and Schepartz, A. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 11735−11739.
15. Kohler, J. J., and Schepartz, A. (2001) *Biochemistry 40*, 130−142.
16. Vinson, C., Myakishev, M., Acharya, A., Mir, A. A., Moll, J. R., and Bonovich, M. (2002) *Mol. Cell. Biol. 22*, 6321−6335.
17. Bosshard, H. R., Dürr, E., Hitz, T., and Jelesarov, I. (2001) *Biochemistry 40*, 3544−3552.
18. König, P., and Richmond, T. J. (1993) *J. Mol. Biol. 233*, 139−154.
19. Berger, C., Jelesarov, I., and Bosshard, H. R. (1996) *Biochemistry 35*, 14984−14991.
20. Leder, L., Berger, C., Bornhauser, S., Wendt, H., Ackermann, F., Jelesarov, I., and Bosshard, H. R. (1995) *Biochemistry 34*, 16509−16518.
21. Kirsch, R. D., and Joly, E. (1998) *Nucleic Acids Res. 26*, 1848−1850.
22. Gill, S. C., and von Hippel, P. (1989) *Anal. Biochem. 182*, 319−326.
23. Brown, T., and Brown, D. J. S. (1991) in *Oligonucleotides and analogues: a practical approach* (Eckstein, F., Ed.) IRL Press, Oxford, U.K.
24. Bernasconi, C. F. (1976) *Relaxation kinetics*, 1st ed., Academic Press, New York.
25. Hinnebusch, A. G. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6442−6446.
26. Hope, I. A., and Struhl, K. (1985) *Cell 43*, 177−188.
27. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) *Science 240*, 1759−1764.
28. Hope, I. A., and Struhl, K. (1987) *EMBO J. 6*, 2781−2784.
29. Lee, W., Mitchell, P., and Tjian, R. (1987) *Cell 49*, 741−752.
30. Sellers, J. W., Vincent, A. C., and Struhl, K. (1990) *Mol. Cell. Biol. 10*, 5077−5086.
31. Kuzmic, P. (1996) *Anal. Biochem. 237*, 260−273.
32. Suckow, M., Schwamborn, K., Kisters-Woike, B., von Wilcken-Bergmann, B., and Muller-Hill, B. (1994) *Nucleic Acids Res. 22*, 4395−4404.
33. Zitzewitz, J. A., Bilsel, O., Luo, J. B., Jones, B. E., and Matthews, C. R. (1995) *Biochemistry 34*, 12812−12819.
34. Patel, L. R., Curran, T., and Kerppola, T. K. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 7360−7364.

35. von Smolukowski, M. (1917) *Z. Phys. Chem. 92*, 129−168.
36. Dürr, E., Jelesarov, I., and Bosshard, H. R. (1999) *Biochemistry 38*, 870−880.
37. Kozlov, A. G., and Lohman, T. M. (2002) *Biochemistry 41*, 6032−6044.
38. Ferreiro, D. U., and de Prat-Gay, G. (2003) *J. Mol. Biol. 331*, 89−99.
39. Kramer, R. H., and Karpen, J. W. (1998) *Nature 395*, 710−713.
40. Iakoucheva, L. M., Walker, R. K., van Houten, B., and Ackerman, E. J. (2002) *Biochemistry 41*, 131−143.
41. Metallo, S. J., and Schepartz, A. (1994) *Chem. Biol. 1*, 143−151.
42. von Hippel, P. H., and Berg, O. G. (1989) *J. Biol. Chem. 264*, 675−678.
43. Shimamoto, N. (1999) *J. Biol. Chem. 274*, 15293−15296.
44. Stanford, N. P., Szczelkun, M. D., Marko, J. F., and Halford, S. E. (2001) *EMBO J. 19*, 6546−6557.
45. Schneider, T. L., and Schepartz, A. (2001) *Biochemistry 40*, 2835−2843.

BI0355793

6. LIST OF REFERENCES

Abkevich, V. I., Gutin, A. M. and Shakhnovich, E. I. (1995) Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. *J Mol Biol 252*, 460-71.

Adler, M. and Scheraga, H. A. (1990) Nonnative isomers of proline-93 and -114 predominate in heat-unfolded ribonuclease A. *Biochemistry 29*, 8211-6.

Agashe, V. R., Shastry, M. C. and Udgaonkar, J. B. (1995) Initial hydrophobic collapse in the folding of barstar. *Nature 377*, 754-7.

Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y. and Fujisawa, T. (2002) Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. *Proc Natl Acad Sci U S A 99*, 1329-34.

Alberts, J., Lewis, Raff, Roberts, Walter (2002) *The Cell*, 4th ed., Garland Science, New York.

Bachmann, A. and Kiefhaber, T. (2001) Apparent two-state tendamistat folding is a sequential process along a defined route. *J Mol Biol 306*, 375-86.

Bai, Y. (1999) Equilibrium amide hydrogen exchange and protein folding kinetics. *J Biomol NMR 15*, 65-70.

Bai, Y. (2003) Hidden intermediates and levinthal paradox in the folding of small proteins. *Biochem Biophys Res Commun 305*, 785-8.

Bai, Y., Sosnick, T. R., Mayne, L. and Englander, S. W. (1995) Protein folding intermediates: native-state hydrogen exchange. *Science 269*, 192-7.

Baldwin, R. L. (1996) On-pathway versus off-pathway folding intermediates. *Fold Des 1*, R1-8.

Baldwin, R. L. and Rose, G. D. (1999) Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem Sci 24*, 77-83.

Baldwin, R. L. and Rose, G. D. (1999i) Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem Sci 24*, 26-33.

Baldwin, R. L. and Rose, G. D. (1999ii) Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem Sci 24*, 77-83.

Beck, C., Cranz, S., Solmaz, M., Roth, M. and Jeltsch, A. (2001) How does a DNA interacting enzyme change its specificity during molecular evolution? A site-directed mutagenesis study at the DNA binding site of the DNA-(adenine-N6)-methyltransferase EcoRV. *Biochemistry 40*, 10956-65.

Beck, C., Siemens, X. and Weaver, D. L. (2001) Diffusion-collision model study of misfolding in a four-helix bundle protein. *Biophys J 81*, 3105-15.

Bennion, B. J. and Daggett, V. (2003) The molecular basis for the chemical denaturation of proteins by urea. *Proc Natl Acad Sci U S A 100*, 5142-7.

Berger, C., Jelesarov, I. and Bosshard, H. R. (1996) Coupled folding and site-specific binding of the GCN4-bZIP transcription factor to the AP-1 and ATF/CREB DNA sites studied by microcalorimetry. *Biochemistry 35*, 14984-91.

Berger, C., Piubelli, L., Haditsch, U. and Bosshard, H. R. (1998) Diffusion-controlled DNA recognition by an unfolded, monomeric bZIP transcription factor. *FEBS Lett 425*, 14-8.

Bernasconi, C. (1976) *Relaxation Kinetics*, 1. Edition ed., Academic Press, New York.

Bosshard, H. R., Durr, E., Hitz, T. and Jelesarov, I. (2001) Energetics of coiled coil folding: the nature of the transition states. *Biochemistry 40*, 3544-52.

Branden, T. (1998) *Introduction to Protein Structure*, Garland Publishing , Inc., New York.

Brandts, J. F., Halvorson, H. R. and Brennan, M. (1975) Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry 14*, 4953-63.

Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L. and Oas, T. G. (1997) The energy landscape of a fast-folding protein mapped by Ala-->Gly substitutions. *Nat Struct Biol 4*, 305-10.

Burton, R. E., Myers, J. K. and Oas, T. G. (1998) Protein folding dynamics: quantitative comparison between theory and experiment. *Biochemistry 37*, 5337-43.

Calloni, G., Taddei, N., Plaxco, K. W., Ramponi, G., Stefani, M. and Chiti, F. (2003) Comparison of the folding processes of distantly related proteins. Importance of hydrophobic content in folding. *J Mol Biol 330*, 577-91.

Capaldi, A. P., Kleanthous, C. and Radford, S. E. (2002) Im7 folding mechanism: misfolding on a path to the native state. *Nat Struct Biol 9*, 209-16.

Capaldi, A. P., Shastry, M. C., Kleanthous, C., Roder, H. and Radford, S. E. (2001) Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat Struct Biol 8*, 68-72.

Chiti, F., Taddei, N., Webster, P., Hamada, D., Fiaschi, T., Ramponi, G. and Dobson, C. M. (1999) Acceleration of the folding of acylphosphatase by stabilization of local secondary structure. *Nat Struct Biol 6*, 380-7.

Chowdhury, S., Zhang, W., Wu, C., Xiong, G. and Duan, Y. (2003) Breaking non-native hydrophobic clusters is the rate-limiting step in the folding of an alanine-based peptide. *Biopolymers 68*, 63-75.

Chu, R. A., Takei, J., Barchi, J. J., Jr. and Bai, Y. (1999) Relationship between the native-state hydrogen exchange and the folding pathways of barnase. *Biochemistry 38*, 14119-24.

Colon, W., Elove, G. A., Wakem, L. P., Sherman, F. and Roder, H. (1996) Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome c folding. *Biochemistry 35*, 5538-49.

Cranz, S., Berger, C., Baici, A., Jelesarov, I. and Bosshard, H. R. (2004) Monomeric and Dimeric bZIP Transcription Factor GCN4 Bind at the Same Rate to Their Target DNA Site. *Biochemistry 43*, 718-27.

Daggett, V. and Fersht, A. R. (2003) Is there a unifying mechanism for protein folding? *Trends Biochem Sci 28*, 18-25.

Dennis, C. A., Videler, H., Pauptit, R. A., Wallis, R., James, R., Moore, G. R. and Kleanthous, C. (1998) A structural comparison of the colicin immunity proteins Im7 and Im9 gives new insights into the molecular determinants of immunity-protein specificity. *Biochem J 333 (Pt 1)*, 183-91.

Dill, K. A. (1985) Theory for the folding and stability of globular proteins. *Biochemistry 24*, 1501-9.

Dill, K. A. (1990) Dominant forces in protein folding. *Biochemistry 29*, 7133-55.

Doyle, R., Simons, K., Qian, H. and Baker, D. (1997) Local interactions and the optimization of protein folding. *Proteins 29*, 282-91.

Eaton, W. A., Thompson, P. A., Chan, C. K., Hage, S. J. and Hofrichter, J. (1996) Fast events in protein folding. *Structure 4*, 1133-9.

Eliezer, D., Yao, J., Dyson, H. J. and Wright, P. E. (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat Struct Biol 5*, 148-55.

Ellenberger, T. E., Brandl, C. J., Struhl, K. and Harrison, S. C. (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell 71*, 1223-37.

Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C. and Radford, S. E. (1999) Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol 286*, 1597-608.
Ferguson, N., Li, W., Capaldi, A. P., Kleanthous, C. and Radford, S. E. (2001) Using chimeric immunity proteins to explore the energy landscape for alpha-helical protein folding. *J Mol Biol 307*, 393-405.
Fernandez, A., Kardos, J. and Goto, Y. (2003) Protein folding: could hydrophobic collapse be coupled with hydrogen-bond formation? *FEBS Lett 536*, 187-92.
Fersht, A. (1998) *Structure and Mechanism in Protein Science*, 1. ed., W.H. Freeman and Company, New York.
Fersht, A. R. (1993) The sixth Datta Lecture. Protein folding and stability: the pathway of folding of barnase. *FEBS Lett 325*, 5-16.
Fersht, A. R., Matouschek, A. and Serrano, L. (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol 224*, 771-82.
Friel, C. T., Capaldi, A. P. and Radford, S. E. (2003) Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: similarities and differences in the folding of homologous proteins. *J Mol Biol 326*, 293-305.
Fujii, Y., Shimizu, T., Toda, T., Yanagida, M. and Hakoshima, T. (2000) Structural basis for the diversity of DNA recognition by bZIP transcription factors. *Nat Struct Biol 7*, 889-93.
Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W., DeMarco, M. L., Daggett, V. and Fersht, A. R. (2003) Unifying features in protein-folding mechanisms. *Proc Natl Acad Sci U S A 100*, 13286-91.
Gill, S. C. and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem 182*, 319-26.
Gong, H., Isom, D. G., Srinivasan, R. and Rose, G. D. (2003) Local secondary structure content predicts folding rates for simple, two-state proteins. *J Mol Biol 327*, 1149-54.
Gorski, S. A., Capaldi, A. P., Kleanthous, C. and Radford, S. E. (2001) Acidic conditions stabilise intermediates populated during the folding of Im7 and Im9. *J Mol Biol 312*, 849-63.
Gorski, S. A., Le Duff, C. S., Capaldi, A. P., Kalverda, A. P., Beddard, G. S., Moore, G. R. and Radford, S. E. (2004) Equilibrium hydrogen exchange reveals extensive hydrogen bonded secondary structure in the on-pathway intermediate of Im7. *J Mol Biol 337*, 183-93.
Govindarajan, S. and Goldstein, R. A. (1995) Optimal local propensities for model proteins. *Proteins 22*, 413-8.
Gowers, D. M. and Halford, S. E. (2003) Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *Embo J 22*, 1410-8.
Gutin, A. M., Abkevich, V. I. and Shakhnovich, E. I. (1995) Is burst hydrophobic collapse necessary for protein folding? *Biochemistry 34*, 3066-76.
Hagerman, P. J. (1990) Pyrimidine 5-methyl groups influence the magnitude of DNA curvature. *Biochemistry 29*, 1980-3.
Halford, S. E. and Szczelkun, M. D. (2002) How to get from A to B: strategies for analysing protein motion on DNA. *Eur Biophys J 31*, 257-67.
Hill, R. B. and DeGrado, W. F. (2000) A polar, solvent-exposed residue can be essential for native protein structure. *Structure Fold Des 8*, 471-9.
Hinnebusch, A. G. (1984) Evidence for translational regulation of the activator of general amino acid control in yeast. *Proc Natl Acad Sci U S A 81*, 6442-6.
Hoang, L., Bedard, S., Krishna, M. M., Lin, Y. and Englander, S. W. (2002) Cytochrome c folding pathway: kinetic native-state hydrogen exchange. *Proc Natl Acad Sci U S A 99*, 12173-8.

Hollenbeck, J. J., McClain, D. L. and Oakley, M. G. (2002) The role of helix stabilizing residues in GCN4 basic region folding and DNA binding. *Protein Sci 11*, 2740-7.
Hollenbeck, J. J. and Oakley, M. G. (2000) GCN4 binds with high affinity to DNA sequences containing a single consensus half-site. *Biochemistry 39*, 6380-9.
Honig, B. (1999) Protein folding: from the levinthal paradox to structure prediction. *J Mol Biol 293*, 283-93.
Huang, C. Y., Getahun, Z., Zhu, Y., Klemke, J. W., DeGrado, W. F. and Gai, F. (2002) Helix formation via conformation diffusion search. *Proc Natl Acad Sci U S A 99*, 2788-93.
Hurst, H. (1996) *Leucine Zippers*, Academic Press Inc., San Diego, CA.
Ibarra-Molero, B., Makhatadze, G. I. and Matthews, C. R. (2001) Mapping the energy surface for the folding reaction of the coiled-coil peptide GCN4-p1. *Biochemistry 40*, 719-31.
Ikai, A. and Tanford, C. (1973) Kinetics of unfolding and refolding of proteins. I. Mathematical analysis. *J Mol Biol 73*, 145-63.
Islam, S. A., Karplus, M. and Weaver, D. L. (2002) Application of the diffusion-collision model to the folding of three-helix bundle proteins. *J Mol Biol 318*, 199-215.
Itzhaki, L. S., Otzen, D. E. and Fersht, A. R. (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol 254*, 260-88.
Jackson, S. E. and Fersht, A. R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry 30*, 10428-35.
Jackson, S. E. and Fersht, A. R. (1991) Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. *Biochemistry 30*, 10436-43.
Jen-Jacobson, L., Engler, L. E. and Jacobson, L. A. (2000) Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure Fold Des 8*, 1015-23.
Jones, S., van Heyningen, P., Berman, H. M. and Thornton, J. M. (1999) Protein-DNA interactions: A structural analysis. *J Mol Biol 287*, 877-96.
Karplus, M. and Weaver, D. L. (1976) Protein-folding dynamics. *Nature 260*, 404-6.
Karplus, M. and Weaver, D. L. (1994) Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci 3*, 650-68.
Keller, W., Konig, P. and Richmond, T. J. (1995) Crystal structure of a bZIP/DNA complex at 2.2 A: determinants of DNA specific recognition. *J Mol Biol 254*, 657-67.
Khandelwal, P., Panchal, S. C., Radha, P. K. and Hosur, R. V. (2001) Solution structure and dynamics of GCN4 cognate DNA: NMR investigations. *Nucleic Acids Res 29*, 499-505.
Khorasanizadeh, S., Peters, I. D. and Roder, H. (1996) Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat Struct Biol 3*, 193-205.
Kim, J., Tzamarias, D., Ellenberger, T., Harrison, S. C. and Struhl, K. (1993) Adaptability at the protein-DNA interface is an important aspect of sequence recognition by bZIP proteins. *Proc Natl Acad Sci U S A 90*, 4513-7.
Kleanthous, C. and Walker, D. (2001) Immunity proteins: enzyme inhibitors that avoid the active site. *Trends Biochem Sci 26*, 624-31.
Kohler, J. J. and Schepartz, A. (2001) Kinetic studies of Fos.Jun.DNA complex formation: DNA binding prior to dimerization. *Biochemistry 40*, 130-42.
Konig, P. and Richmond, T. J. (1993) The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J Mol Biol 233*, 139-54.
Koonin, E. V., Wolf, Y. I. and Karev, G. P. (2002) The structure of the protein universe and genome evolution. *Nature 420*, 218-23.

Kraulis, P. J. (1991) MOLSCRIPT- a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog. 24*, 946-950.
Lacroix, E., Viguera, A. R. and Serrano, L. (1998) Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J Mol Biol 284*, 173-91.
Levine, M. and Tjian, R. (2003) Transcription regulation and animal diversity. *Nature 424*, 147-51.
Levinthal, C. (1968) Are there pathways in protein folding? *J. Chim. Phys. 65*, 44-45.
Lopez-Hernandez, E., Cronet, P., Serrano, L. and Munoz, V. (1997) Folding kinetics of Che Y mutants with enhanced native alpha-helix propensities. *J Mol Biol 266*, 610-20.
Luscombe, N. M., Austin, S. E., Berman, H. M. and Thornton, J. M. (2000) An overview of the structures of protein-DNA complexes. *Genome Biol 1*, REVIEWS001.
Maher, L. J., 3rd. (1998) Mechanisms of DNA bending. *Curr Opin Chem Biol 2*, 688-94.
Matouschek, A., Serrano, L. and Fersht, A. R. (1992) The folding of an enzyme. IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. *J Mol Biol 224*, 819-35.
Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S., Freund, S. M., Alonso, D. O., Daggett, V. and Fersht, A. R. (2003) The complete folding pathway of a protein from nanoseconds to microseconds. *Nature 421*, 863-7.
Meng, F. G., Zeng, X., Hong, Y. K. and Zhou, H. M. (2001) Dissociation and unfolding of GCN4 leucine zipper in the presence of sodium dodecyl sulfate. *Biochimie 83*, 953-6.
Merrit, E. A., Bacon, D.J. (1997) Raster 3D: photorealistic molecular graphics. *Methods Enzymol. 277*, 505-524.
Miletich, I. and Sharpe, P. T. (2003) Normal and abnormal dental development. *Hum Mol Genet 12 Spec No 1*, R69-73.
Miller, M., Shuman, J. D., Sebastian, T., Dauter, Z. and Johnson, P. F. (2003) Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. *J Biol Chem 278*, 15178-84.
Misteli, T. (2001) Protein dynamics: implications for nuclear architecture and gene expression. *Science 291*, 843-7.
Moss, T. (2001) *DNA-Protein Interactions*, Vol. 148., Second Edition ed., Humana Press, Totowa.
Munoz, V., Cronet, P., Lopez-Hernandez, E. and Serrano, L. (1996) Analysis of the effect of local interactions on protein stability. *Fold Des 1*, 167-78.
Munoz, V., Lopez, E. M., Jager, M. and Serrano, L. (1994) Kinetic characterization of the chemotactic protein from Escherichia coli, CheY. Kinetic analysis of the inverse hydrophobic effect. *Biochemistry 33*, 5858-66.
Munoz, V. and Serrano, L. (1994) Elucidating the folding problem of helical peptides using empirical parameters. *Nat Struct Biol 1*, 399-409.
Munoz, V. and Serrano, L. (1995) Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J Mol Biol 245*, 275-96.
Munoz, V. and Serrano, L. (1997) Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers 41*, 495-509.
Myers, J. K. and Oas, T. G. (1999) Reinterpretation of GCN4-p1 folding kinetics: partial helix formation precedes dimerization in coiled coil folding. *J Mol Biol 289*, 205-9.
Myers, J. K. and Oas, T. G. (2001) Preorganized secondary structure as an important determinant of fast protein folding. *Nat Struct Biol 8*, 552-8.
Natarajan, K., Meyer, M. R., Jackson, B. M., Slade, D., Roberts, C., Hinnebusch, A. G. and Marton, M. J. (2001) Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. *Mol Cell Biol 21*, 4347-68.

Newman, J. R. and Keating, A. E. (2003) Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science 300*, 2097-101.
Nolting, B. and Andert, K. (2000) Mechanism of protein folding. *Proteins 41*, 288-98.
Nolting, B., Golbik, R., Neira, J. L., Soler-Gonzalez, A. S., Schreiber, G. and Fersht, A. R. (1997) The folding pathway of a protein at high resolution from microseconds to seconds. *Proc Natl Acad Sci U S A 94*, 826-30.
Northey, J. G., Di Nardo, A. A. and Davidson, A. R. (2002) Hydrophobic core packing in the SH3 domain folding transition state. *Nat Struct Biol 9*, 126-30.
Oda, M. and Nakamura, H. (2000) Thermodynamic and kinetic analyses for understanding sequence-specific DNA recognition. *Genes Cells 5*, 319-26.
Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z. and Wolynes, P. G. (1996) Protein folding funnels: the nature of the transition state ensemble. *Fold Des 1*, 441-50.
Osborne, M. J., Breeze, A. L., Lian, L. Y., Reilly, A., James, R., Kleanthous, C. and Moore, G. R. (1996) Three-dimensional solution structure and 13C nuclear magnetic resonance assignments of the colicin E9 immunity protein Im9. *Biochemistry 35*, 9505-12.
Otzen, D. E., Itzhaki, L. S., elMasry, N. F., Jackson, S. E. and Fersht, A. R. (1994) Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proceedings Of The National Academy Of Sciences Of The United States Of America 91*, 10422-10425.
Paci, E., Friel, C. T., Lindorff-Larsen, K., Radford, S. E., Karplus, M. and Vendruscolo, M. (2004) Comparison of the transition state ensembles for folding of Im7 and Im9 determined using all-atom molecular dynamics simulations with phi value restraints. *Proteins 54*, 513-25.
Packham, E. A. and Brook, J. D. (2003) T-box genes in human disorders. *Hum Mol Genet 12 Spec No 1*, R37-44.
Pain, R. H. (1994) *Mechanisms of Protein Folding*, 1. ed., Oxford Universtiy Press, Oxford.
Pakula, A. A. and Sauer, R. T. (1990) Reverse hydrophobic effects relieved by amino-acid substitutions at a protein surface. *Nature 344*, 363-4.
Perl, D., Welker, C., Schindler, T., Schroder, K., Marahiel, M. A., Jaenicke, R. and Schmid, F. X. (1998) Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat Struct Biol 5*, 229-35.
Phillips, S. E. (1994) Built by association: structure and function of helix-loop-helix DNA-binding proteins. *Structure 2*, 1-4.
Plaxco, K. W., Simons, K. T. and Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol 277*, 985-94.
Poso, D., Sessions, R. B., Lorch, M. and Clarke, A. R. (2000) Progressive stabilization of intermediate and transition states in protein folding reactions by introducing surface hydrophobic residues. *J Biol Chem 275*, 35723-6.
Privalov, P. L. and Gill, S. J. (1988) Stability of protein structure and hydrophobic interaction. *Adv Protein Chem 39*, 191-234.
Richards, F. M. and Lim, W. A. (1993) An analysis of packing in the protein folding problem. *Q Rev Biophys 26*, 423-98.
Robinson, H., Gao, Y. G., McCrary, B. S., Edmondson, S. P., Shriver, J. W. and Wang, A. H. (1998) The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature 392*, 202-5.
Rose, G. D. and Roy, S. (1980) Hydrophobic basis of packing in globular proteins. *Proc Natl Acad Sci U S A 77*, 4643-7.
Sabelko, J., Ervin, J. and Gruebele, M. (1999) Observation of strange kinetics in protein folding. *Proc Natl Acad Sci U S A 96*, 6031-6.
Sadqi, M., Lapidus, L. J. and Munoz, V. (2003) How fast is protein hydrophobic collapse? *Proc Natl Acad Sci U S A 100*, 12117-22.

Sanchez, I. E. and Kiefhaber, T. (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J Mol Biol 325*, 367-76.
Sanchez, I. E. and Kiefhaber, T. (2003) Non-linear rate-equilibrium free energy relationships and Hammond behavior in protein folding. *Biophys Chem 100*, 397-407.
Schindler, T., Herrler, M., Marahiel, M. A. and Schmid, F. X. (1995) Extremely rapid protein folding in the absence of intermediates. *Nat Struct Biol 2*, 663-73.
Serrano, L., Matouschek, A. and Fersht, A. R. (1992) The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J Mol Biol 224*, 805-18.
Sosnick, T. R., Jackson, S., Wilk, R. R., Englander, S. W. and DeGrado, W. F. (1996) The role of helix formation in the folding of a fully alpha-helical coiled coil. *Proteins 24*, 427-32.
Srinivasan, R. and Rose, G. D. (1999) A physical basis for protein secondary structure. *Proc Natl Acad Sci U S A 96*, 14258-63.
Stanford, N. P., Szczelkun, M. D., Marko, J. F. and Halford, S. E. (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites. *Embo J 19*, 6546-57.
Stanojevic, D. and Verdine, G. L. (1995) Deconstruction of GCN4/GCRE into a monomeric peptide-DNA complex. *Nat Struct Biol 2*, 450-7.
Stryer, L. (1990) *Biochemistry*, 3. ed., W.H. Freeman and Company, New York.
Suckow, M. and Hollenberg, C. P. (1998) The activation specificities of wild-type and mutant Gcn4p in vivo can be different from the DNA binding specificities of the corresponding bZip peptides in vitro. *J Mol Biol 276*, 887-902.
Suckow, M., Madan, A., Kisters-Woike, B., von Wilcken-Bergmann, B. and Muller-Hill, B. (1994) Creating new DNA binding specificities in the yeast transcriptional activator GCN4 by combining selected amino acid substitutions. *Nucleic Acids Res 22*, 2198-208.
Suckow, M., Schwamborn, K., Kisters-Woike, B., von Wilcken-Bergmann, B. and Muller-Hill, B. (1994) Replacement of invariant bZip residues within the basic region of the yeast transcriptional activator GCN4 can change its DNA binding specificity. *Nucleic Acids Res 22*, 4395-404.
Suckow, M., von Wilcken-Bergmann, B. and Muller-Hill, B. (1993) The DNA binding specificity of the basic region of the yeast transcriptional activator GCN4 can be changed by substitution of a single amino acid. *Nucleic Acids Res 21*, 2081-6.
Suckow, M., von Wilcken-Bergmann, B. and Muller-Hill, B. (1993) Identification of three residues in the basic regions of the bZIP proteins GCN4, C/EBP and TAF-1 that are involved in specific DNA binding. *Embo J 12*, 1193-200.
Taddei, N., Chiti, F., Fiaschi, T., Bucciantini, M., Capanni, C., Stefani, M., Serrano, L., Dobson, C. M. and Ramponi, G. (2000) Stabilisation of alpha-helices by site-directed mutagenesis reveals the importance of secondary structure in the transition state for acylphosphatase folding. *J Mol Biol 300*, 633-47.
Takeda, Y., Ross, P. D. and Mudd, C. P. (1992) Thermodynamics of Cro protein-DNA interactions. *Proc Natl Acad Sci U S A 89*, 8180-4.
Takei, J., Pei, W., Vu, D. and Bai, Y. (2002) Populating partially unfolded forms by hydrogen exchange-directed protein engineering. *Biochemistry 41*, 12308-12.
Travers, A. (1992) DAN confomation and configuration in protein-DNA complexes. *Current Opinion in Structural Biology 2*, 71-77.
Uil, T. G., Haisma, H. J. and Rots, M. G. (2003) Therapeutic modulation of endogenous gene function by agents with designed DNA-sequence specificities. *Nucleic Acids Res 31*, 6064-78.
Unger, R. and Moult, J. (1996) Local interactions dominate folding in a simple protein model. *J Mol Biol 259*, 988-94.

Uversky, V. N. (2002) Cracking the folding code. Why do some proteins adopt partially folded conformations, whereas other don't? *FEBS Lett 514*, 181-3.
Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., et al. (2001) The sequence of the human genome. *Science 291*, 1304-51.
Viguera, A. R., Vega, C. and Serrano, L. (2002) Unspecific hydrophobic stabilization of folding transition states. *Proc Natl Acad Sci U S A 99*, 5349-54.
Viguera, A. R., Villegas, V., Aviles, F. X. and Serrano, L. (1997) Favourable native-like helical local interactions can accelerate protein folding. *Fold Des 2*, 23-33.
Villegas, V., Viguera, A. R., Aviles, F. X. and Serrano, L. (1996) Stabilization of proteins by rational design of alpha-helix stability using helix/coil transition theory. *Fold Des 1*, 29-34.
Villegas, V. V., Viguera, A. R., Aviles, F. X. and Serrano, L. (1995) Stabilization of proteins by rational design of alpha-helix stability using helix/coil transition theory. *Fold Des 1*, 29-34.
Vinson, C., Myakishev, M., Acharya, A., Mir, A. A., Moll, J. R. and Bonovich, M. (2002) Classification of human B-ZIP proteins based on dimerization properties. *Mol Cell Biol 22*, 6321-35.
Waddick, K. G. and Uckun, F. M. (1998) Innovative treatment programs against cancer. I. Ras oncoprotein as a molecular target. *Biochem Pharmacol 56*, 1411-26.
Waddick, K. G. and Uckun, F. M. (1999) Innovative treatment programs against cancer: II. Nuclear factor-kappaB (NF-kappaB) as a molecular target. *Biochem Pharmacol 57*, 9-17.
Wang, X., Cao, W., Cao, A. and Lai, L. (2003) Thermodynamic characterization of the folding coupled DNA binding by the monomeric transcription activator GCN4 peptide. *Biophys J 84*, 1867-75.
Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C. and Struhl, K. (1990) Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature 347*, 575-8.
Wolberger, C. (1999) Multiprotein-DNA complexes in transcriptional regulation. *Annu Rev Biophys Biomol Struct 28*, 29-56.
Wu, Y. and Marsh, J. W. (2003) Gene transcription in HIV infection. *Microbes Infect 5*, 1023-7.
Zhou, Y. and Karplus, M. (1999) Interpreting the folding kinetics of helical proteins. *Nature 401*, 400-3.
Zhu, Y., Alonso, D. O., Maki, K., Huang, C. Y., Lahr, S. J., Daggett, V., Roder, H., DeGrado, W. F. and Gai, F. (2003) Ultrafast folding of {alpha}3D: A de novo designed three-helix bundle protein. *Proc Natl Acad Sci U S A*.
Zitzewitz, J. A., Bilsel, O., Luo, J., Jones, B. E. and Matthews, C. R. (1995) Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry 34*, 12812-9.

7. ACKNOWLEDGMENTS

First I would like to thank Prof. Dr. Bosshard for giving me an opportunity to work with him and his group. It was a pleasure.

I would like to thank all of the members of the Bosshard group, which gave me encouragement and criticism, whichever seemed more appropriate at the time. Although it was not always obvious, I do appreciate both.

I am also grateful to Prof. Dr. Sheena Radford in Leeds, UK. She gave me the chance to spend one year in her vibrant lab. It was inspiring.

The Radford Group in Leeds also deserves special thanks for teaching me so much about science and putting up with my home sickness. I wanted to feel terrible, but they wouldn't let me.

Most importantly, I would like to thank my family and all of my friends. Each and everyone supported me in the best way they could. A special mention goes to Eva Sanchez-Cobos and Clemens Stilling for giving me shelter when I became homeless in Leeds.

8. CURRICULUM VITAE

Personal data

Name	Cranz
First name	Susanne
Date of birth	16th of March 1976 in
Place of birth	Ingelheim am Rhein
Citizen of	Germany

Education

1995	Abitur, Albert Schweitzer Gymnasium, Leonberg, Germany Recipient of the Karl-von-Frisch Award (vdbiol)
1996	Undergraduate studies in biology at the Justus-Liebig-Universität, Giessen
2000	Diploma Thesis at the Department of Biochemistry: **"Characterisation of the DNA-Binding Domain of the Methytransferase M. *Eco*RV by Site-Directed Mutagenesis"** (PD Dr. Albert Jeltsch)
2000-2004	PhD research in the laboratory of Prof. Dr. H. R. Bosshard, Zurich, Switzerland
2002-2003	PhD research with Prof. Dr. S. E. Radford. Leeds, UK (Marie Curie Host Fellowship)
2004	Dissertation: **"Helix Stability and Hydrophobicity in the Folding of Immunity Protein Im9 and Kinetics of DNA-Binding to Transcription Factor GCN4"**